

# A PROBABILISTIC MODEL FOR THE DISTRIBUTION OF RANKS OF ELLIPTIC CURVES OVER $\mathbb{Q}$

ÁLVARO LOZANO-ROBLEDO

ABSTRACT. In this article, we propose a new probabilistic model for the distribution of ranks of elliptic curves in families of fixed Selmer rank, and compare the predictions with previous results, and with the databases of curves over the rationals that we have at our disposal. In addition, we document a phenomenon we refer to as *Selmer bias* that seems to play an important role in the data and in our models.

## 1. INTRODUCTION

Let  $E/\mathbb{Q}$  be an elliptic curve. The Mordell-Weil theorem states that the group  $E(\mathbb{Q})$  of rational points on  $E$  is finitely generated and, therefore, we have an isomorphism

$$E(\mathbb{Q}) \cong E(\mathbb{Q})_{\text{tors}} \oplus \mathbb{Z}^{R_E},$$

where  $E(\mathbb{Q})_{\text{tors}}$  is the (finite) subgroup of points of finite order, and  $R_E = \text{rank}(E(\mathbb{Q})) \geq 0$  is the rank of the elliptic curve. The torsion subgroups that arise over  $\mathbb{Q}$  are well understood: Mazur's theorem settles what groups are possible ([19], [20]), the parametrization of the corresponding modular curves are known ([18]), and we know the distribution of elliptic curves with a prescribed torsion subgroup ([13]) as a function of the height of the curve. However, the distribution of ranks of elliptic curves is unknown. Several conjectures can be found in the literature (e.g., on the average rank, see [22]), and also some heuristic models ([27], [21]), but the basic questions about the distribution of the ranks remain unanswered. For instance, it is not known whether the rank can be arbitrarily large (currently, the largest rank known is 28, due to Noam Elkies - see [10] for Elkies' example, and other current records).

In this article, we propose a new probabilistic model for the distribution of ranks of elliptic curves (in families of fixed 2-Selmer rank) and explore its possible consequences. In addition, we document a phenomenon we refer to as *Selmer bias* that seems to play an important role in the data and in our models. We use the largest database of elliptic curves at our disposal ([1], which we will refer to as the BHKSSW database) in order to test our model and to make predictions. We have also computed additional data for one million curves in the family of elliptic curves with  $j = 1728$  that are also used to test the model.

In our model, we fix  $r \geq 0$ , and we study the probability that an elliptic curve  $E/\mathbb{Q}$  of naive height  $X$  has rank  $r \geq 0$  (the naive height will be defined in Section 3). We will assume throughout that the 2-primary part of the Tate-Shafarevich group of an elliptic curve  $E/\mathbb{Q}$ , denoted by  $\text{III}(E/\mathbb{Q})[2^\infty]$ , is finite (therefore, by the existence of the Cassels-Tate pairing, its size is a square, which in turn implies a number of parity considerations). In particular, if we define the (2-)Selmer rank of an

elliptic curve by  $\text{selrank}(E(\mathbb{Q})) = \dim_{\mathbb{F}_2} \text{Sel}_2(E/\mathbb{Q}) - \dim_{\mathbb{F}_2}(E[2])$ , then the Selmer rank and the rank agree modulo 2. Let us fix some notation:

- For fixed  $n, r \geq 0$ , and for any  $1 \leq X_1 \leq X_2$ , let  $\mathcal{E}([X_1, X_2])$ ,  $\mathcal{S}_n([X_1, X_2])$ , and  $\mathcal{R}_r([X_1, X_2])$  be, respectively, the sets of all elliptic curves, all curves with Selmer rank  $n$ , and curves of rank  $r$ , with naive height in the interval  $[X_1, X_2]$ .
- We will denote the set of elliptic curves of height exactly  $X$  by  $\mathcal{E}^X = \mathcal{E}([X, X])$ , and we will write  $\mathcal{E}(X) = \mathcal{E}([1, X])$  for the set of all elliptic curves up to height  $X$ . We define similarly  $\mathcal{S}_n^X$ ,  $\mathcal{S}_n(X)$ , and  $\mathcal{R}_r^X$ ,  $\mathcal{R}_r(X)$ , for each  $n, r \geq 0$ .
- If  $\mathcal{C} \subseteq \mathcal{E}$  is a set of elliptic curves (say  $\mathcal{C} = \mathcal{E}$ ,  $\mathcal{S}_n$ ,  $\mathcal{R}_r$ , or  $\mathcal{R}_r \cap \mathcal{S}_n$ ), then we write  $\pi_{\mathcal{C}}(X)$  for  $\#\mathcal{C}([1, X])$ , i.e.,  $\pi_{\mathcal{C}}$  is the counting function of elliptic curves in  $\mathcal{C}$  up to height  $X$ .

For a fixed rank  $r \geq 0$  and a height  $X$  such that the set  $\mathcal{E}^X$  is non-empty, we are interested in the probability that an elliptic curve of height  $X$  belongs to  $\mathcal{R}_r^X$ , that is,  $\text{Prob}(E \in \mathcal{R}_r^X) = \#\mathcal{R}_r^X / \#\mathcal{E}^X$ . Our model is based on the probability formula:

$$\text{Prob}(E \in \mathcal{R}_r^X) = \sum_{j \geq 0} \text{Prob}(E \in \mathcal{S}_{r+2j}^X) \cdot \text{Prob}(E \in \mathcal{R}_r^X \mid E \in \mathcal{S}_{r+2j}^X),$$

where  $\text{Prob}(E \in \mathcal{S}_{r+2j}^X) = \#\mathcal{S}_{r+2j}^X / \#\mathcal{E}^X$ , and we define the conditional probability  $\text{Prob}(E \in \mathcal{R}_r^X \mid E \in \mathcal{S}_{r+2j}^X)$  as 0 if  $\mathcal{S}_{r+2j}^X$  is empty, and by  $\#\mathcal{R}_r^X \cap \mathcal{S}_{r+2j}^X / \#\mathcal{S}_{r+2j}^X$  otherwise. In Section 3 we will discuss the known results about the number of elliptic curves up to height  $X$ . In Sections 4 and 5, and for fixed  $n \geq 0$  and  $X \geq 1$ , we state two probabilistic hypotheses,  $H_A$  and  $H_B$  stated in Hypotheses 4.1 and 5.3 respectively, that our model will be rest on, and which we summarize next:

- ( $H_A$ ) **Hypothesis A:** let  $Y_{\text{Sel},n,X} : \mathcal{E}^X \rightarrow \{0, 1\}$  be the function such that  $Y_{\text{Sel},n,X}(E/\mathbb{Q}) = 1$  if  $E \in \mathcal{E}^X$  has 2-Selmer rank  $n$ , and  $Y_{\text{Sel},n,X}(E) = 0$  otherwise. Then, there is a function  $\theta_n(X)$  such that  $Y_{\text{Sel},n,X}$  behaves as a random variable with Bernoulli distribution  $B(1, \theta_n(X))$ . In particular, this implies that the expected value  $\mathbb{E}(Y_{\text{Sel},n,X})$  is  $\text{Prob}(E \in \mathcal{S}_n^X) = \theta_n(X)$ .
- ( $H_B$ ) **Hypothesis B:** let  $n$  be even (the odd case is slightly different, see 5.3), let  $\text{Sel}_n^X$  be a set formed by  $\mathbb{F}_2$ -generators of  $\text{Sel}_2(E/\mathbb{Q})$ , for each  $E \in \mathcal{S}_n^X$ , and let  $Y_{\text{Hasse},n,X} : \text{Sel}_n^X \rightarrow \{0, 1\}$  be the function such that  $Y_{\text{Hasse},n,X}(s_E) = 1$  whenever  $s_E$  is trivial in  $\text{III}(E/\mathbb{Q})[2]$ , and  $Y_{\text{Hasse},n,X} = 0$  otherwise. Then, there is a function  $\rho_n(X)$  such that  $Y_{\text{Hasse},n,X}$  behaves as a random variable with Bernoulli distribution  $B(1, \rho_n(X))$ . From this distribution, we shall recover the conditional probability  $\text{Prob}(E \in \mathcal{R}_r^X \mid E \in \mathcal{S}_n^X)$  for any  $0 \leq r \leq n$  with  $n \equiv r \pmod{2}$  (see Corollary 5.14).

After taking all the available data under consideration (mainly [1]), we formulate the following conjecture (which is Conjecture 4.5 plus 5.19):

**Conjecture 1.1.** *Hypotheses A and B hold for elliptic curves over  $\mathbb{Q}$ . Moreover, there are constants  $C_n$ ,  $D_n$ ,  $e_n$ ,  $f_n$ , for each  $n \geq 1$ , such that*

$$\theta_n(X) = \frac{s_n}{1 + C_n X^{-e_n}}, \quad \text{and} \quad \rho_n(X) = \frac{D_n}{X^{f_n}},$$

where the limit values  $s_n$  of  $\theta_n(X)$  are those given by a conjecture of Poonen and Rains, and all constants are positive except  $C_1 < 0$ .

The data suggest that the values of the constants of Conjecture 1.1, for  $n = 1, \dots, 5$ , are as given in Tables 5 and 10, and the limit values  $s_n$  are discussed in Section 4 (as in [22]). In our main

Theorems 6.1 and 7.2, under the assumption of  $H_A$  and  $H_B$ , we provide formulas for  $\pi_{\mathcal{R}_r \cap \mathcal{S}_n}(X)$ , i.e., the number of elliptic curves of rank  $r$  and Selmer rank  $n$  up to height  $X$ , and also for the contribution to the average rank coming from elliptic curves of Selmer rank  $n$ .

**Theorem 1.2** (also Theorem 6.1). *Let  $X, r \geq 0, j \geq 0$  be fixed, such that  $n(j) = r + 2j \geq 1$ . If we assume Conjecture 1.1, then*

$$\begin{aligned} \pi_{\mathcal{R}_r \cap \mathcal{S}_{n(j)}}(X) &= \frac{5\kappa}{6} \binom{\lfloor \frac{r}{2} \rfloor + j}{j} \int_0^X \frac{\theta_{n(j)}(H)}{H^{1/6}} \cdot \mathbb{E}_{\lfloor \frac{r}{2} \rfloor, j}^{n(j)}(H) dH + O(X^{1/2}) \\ &= \frac{5\kappa}{6} \binom{\lfloor \frac{r}{2} \rfloor + j}{j} \int_0^X \frac{s_{n(j)} \cdot \mathbb{E}_{\lfloor \frac{r}{2} \rfloor, j}^{n(j)}(H)}{(1 + C_{n(j)} H^{-e_{n(j)}}) \cdot H^{1/6}} dH + O(X^{1/2}), \end{aligned}$$

where  $\kappa = 2^{4/3} \cdot (\zeta(10) \cdot 3^{3/2})^{-1}$ , and  $\mathbb{E}_{\lfloor \frac{r}{2} \rfloor, j}^{n(j)}(H)$  is the expected value defined in Remark 5.12.

In Corollary 6.4 we specialize the formulas of  $\pi_{\mathcal{R}_r \cap \mathcal{S}_{n(j)}}(X)$  for  $0 \leq r \leq n \leq 5$  (see also Table 13). Using our formulas, we have computed approximations of  $\pi_{\mathcal{R}_r}(X)$  for  $1 \leq r \leq 5$  in the range  $[0, 2.7 \cdot 10^{10}]$ , and plotted them in Figures 16 and 17. The error in our approximations is less than 0.7% in this range (see Table 14).

Our second theorem gives conjectural formulas for the contribution to the average rank of elliptic curves coming from elliptic curves of each Selmer rank  $n \geq 1$ . Then, the contributions are added up to estimate the behavior of the average rank.

**Theorem 1.3** (also Theorem 7.2 and Corollary 7.4). *Assume  $H_A$  and  $H_B$ , let  $n \geq 1$  be fixed. Then, the expected value of  $\text{AvgRank}_{\mathcal{S}_n}(X) = \frac{1}{\pi_{\mathcal{E}}(X)} \cdot \sum_{E \in \mathcal{S}_n(X)} \text{rank}(E(\mathbb{Q}))$  is given by*

$$\frac{5\kappa}{6\pi_{\mathcal{E}}(X)} \cdot \int_0^X \frac{\theta_n(H)}{H^{1/6}} \left( (n \bmod 2) + 2 \left\lfloor \frac{n}{2} \right\rfloor \rho_n(H) \right) dH + \theta_n(X) \cdot O(X^{-1/3}).$$

Moreover, the error in approximating  $\text{AvgRank}_{\mathcal{S}_n}(X)$  by its expected value is given by

$$\sqrt{\frac{5\kappa \lfloor n/2 \rfloor}{6\pi_{\mathcal{E}}(X)^2} \int_0^X \frac{\theta_n(H)}{H^{1/6}} (\rho_n(H)(1 - \rho_n(H)) + (\lfloor n/2 \rfloor - 1)C_{1,1}^n(H)) dH + O(X^{-7/6})},$$

where  $C_{1,1}^n(X)$  is the covariance function defined in Proposition 5.8. Further, there are constants  $\tau_n$  such that the expected value of  $\text{AvgRank}_{\mathcal{E}}(X) = \sum_{n=1}^{\infty} \text{AvgRank}_{\mathcal{S}_n}(X)$  is given by

$$\sum_{n=1}^{\infty} s_n \cdot \left( \frac{\tau_n}{X^{5/6}} + \sum_{m=0}^{\infty} \left( \frac{(n \bmod 2)(-C_n)^m}{1 - (6/5)m e_n} + X^{-f_n} \frac{2 \lfloor \frac{n}{2} \rfloor D_n(-C_n)^m}{1 - (6/5)(f_n + m e_n)} \right) X^{-m e_n} \right) + O(X^{-1/3}).$$

In particular,

$$\lim_{X \rightarrow \infty} \text{AvgRank}_{\mathcal{E}}(X) = \sum_{k=0}^{\infty} s_{2k+1} = \frac{1}{2},$$

with standard error going to 0 as  $X \rightarrow \infty$ .

In particular, Theorem 1.3 says that our assumptions imply the so-called “50% – 50% conjecture” (see Conjecture 7.1) and, moreover, it predicts not only the 1/2 limit of the average rank, but also a rate of convergence to said limit. We have used our formulas to compute an approximation of

$\text{AvgRank}_{\mathcal{E}}(X) \approx \sum_{n=1}^5 \text{AvgRank}_{\mathcal{S}_n}(X)$  in the range  $[0, 2.7 \cdot 10^{10}]$  and plotted it in Figure 18. The error in our approximation of  $\text{AvgRank}_{\mathcal{E}}(2.7 \cdot 10^{10})$  is 0.0523% of the actual value (see Remark 7.5). In Table 1, and under the assumption of Conjecture 1.1, we have computed approximate values of  $\text{AvgRank}_{\mathcal{E}}(X) \approx \sum_{n=1}^5 \text{AvgRank}_{\mathcal{S}_n}(X)$  using numerical integration of the formulas of Theorem 1.3.

$X$	$\sum_{n=1}^5 \text{AvgRank}_{\mathcal{S}_n}(X)$	$X$	$\sum_{n=1}^5 \text{AvgRank}_{\mathcal{S}_n}(X)$
$10^{10}$	0.905665	$10^{50}$	0.548880
$10^{15}$	0.846828	$10^{75}$	0.512531
$10^{20}$	0.766868	$10^{100}$	0.503256
$10^{30}$	0.649901	$10^{150}$	0.500215
$10^{40}$	0.585108	$10^{200}$	0.500006

TABLE 1. Conjectural approximate values of  $\sum_{n=1}^5 \text{AvgRank}_{\mathcal{S}_n}(X)$  obtained using numerical integration of the formulas of Theorem 7.2. The integration was done with SageMath, which reports an absolute error in the numerical integration less than  $4 \cdot 10^{-7}$  in all cases. By Theorem 7.2, the limit should be  $s_1 + s_3 + s_5 \approx 0.49999965$ .

Our Hypothesis A also predicts a formula for the average 2-Selmer rank of an elliptic curve.

**Theorem 1.4** (Also Prop. 4.10). *Let  $\text{AvgSelRank}(X)$  be defined by*

$$\text{AvgSelRank}(X) = \frac{1}{\pi_{\mathcal{E}}(X)} \sum_{E \in \mathcal{E}(X)} \text{selrank}(E(\mathbb{Q})).$$

*If we assume  $H_A$  and we assume that  $0 \leq \theta_n(X) \leq s_n$  for all  $n \geq 2$  and all  $X > 0$ , then the expected value of the average Selmer rank is given by*

$$\mathbb{E}(\text{AvgSelRank}(X)) = \frac{5/6}{X^{5/6}} \int_0^X \frac{\sum_{n \geq 1} n \cdot \theta_n(H)}{H^{1/6}} dH + O(X^{-1/3}).$$

*Moreover,  $\lim_{X \rightarrow \infty} \mathbb{E}(\text{AvgSelRank}(X)) = \sum_{n \geq 1} n \cdot s_n = 1.26449978 \dots$*

Finally, as mentioned earlier, a question on *Selmer rank bias* arises in our work:

**Question 1.5.** Does the expected value of  $Y_{\text{Hasse}, n, X}(s_E)$  depend on  $n$ ? In other words, does the probability that  $s_E \in \text{Sel}_2(E/\mathbb{Q})$  is globally solvable depend on  $n = \text{selrank}(E(\mathbb{Q}))$ ?

The answer, surprisingly, seems to be that the probability does depend not only on the parity of  $n$ , but also on the value of  $n$  itself (see Fig. 9). For instance, the data suggest that an element of  $\text{Sel}_2(E/\mathbb{Q})$  is significantly more likely to be globally solvable for  $n = 5$  than for  $n = 3$ . However, the probabilities for  $n = 2$  and  $n = 4$  are quite similar (but they do not behave identically).

**Remark 1.6.** In this article we work with elliptic curves over  $\mathbb{Q}$  and 2-Selmer groups because the database we have to test our models ([1]) only contains 2-Selmer information. However, the same probabilistic model could be derived for  $p$ -Selmer groups over a global field  $K$ .

**Acknowledgements.** The author would like to thank Jennifer Balakrishnan, Keith Conrad, Harris Daniels, Ari Shnidman, Drew Sutherland, and John Voight for their helpful comments and suggestions.

## 2. NOTATION AND PROBABILITY

$\mathcal{E}$	Set of elliptic curves over $\mathbb{Q}$ up to isomorphism	§3
$\text{selrank}(E(\mathbb{Q}))$	2-Selmer rank, equal to $\dim_{\mathbb{F}_2} \text{Sel}_2(E/\mathbb{Q}) - \dim_{\mathbb{F}_2}(E[2])$	§1, 4
$\mathcal{S}_n$	For $n \geq 0$ , curves $E \in \mathcal{E}$ with $\text{selrank}(E(\mathbb{Q})) = n$	§4
$\mathcal{R}_r$	For $r \geq 0$ , curves $E \in \mathcal{E}$ with $\text{rank}(E(\mathbb{Q})) = r$	§6
$\text{ht}(E)$	The naive height of an elliptic curve	§3
$\mathcal{C}(X)$	For $X \geq 0$ , curves in $\mathcal{C} = \mathcal{E}, \mathcal{R}_r$ , or $\mathcal{S}_n$ , with (naive) height $\leq X$	§3, 4, 6
$\mathcal{C}(I)$	For an interval $I$ , curves in $\mathcal{C}$ with height in $I$	§3, 4, 6
$\mathcal{C}^X$	For $X \geq 0$ , curves in $\mathcal{C}$ with height exactly $X$	§3, 4, 6
$\pi_{\mathcal{C}}(X)$	For a set $\mathcal{C} \subseteq \mathcal{E}$ , the size of $\mathcal{C} \cap \mathcal{E}(X)$ , where $\mathcal{C} = \mathcal{E}, \mathcal{R}_r$ , or $\mathcal{S}_n$	§3, 4, 6
$\pi_{\mathcal{C}}(I)$	For a set $\mathcal{C} \subseteq \mathcal{E}$ and an interval $I$ , the size of $\mathcal{C} \cap \mathcal{E}(I)$	§3, 4, 5
$\kappa$	Constant equal to $2^{4/3} \cdot (\zeta(10) \cdot 3^{3/2})^{-1} \approx 0.484462004349 \dots$	Thm. 3.1
$s_n$	$\lim_{X \rightarrow \infty} \pi_{\mathcal{S}_n}(X)/\pi_{\mathcal{E}}(X)$ , given by a conjectural formula by [22]	§4
$B(m, p)$	Binomial distribution with $m$ experiments and probability $p$	§4
$Y_{\text{Sel}, n, X}(E/\mathbb{Q})$	Random variable with value 1 if $\text{selrank}(E(\mathbb{Q})) = n$ , and 0 otherwise	Hyp. 4.1
$\theta_n(X)$	The function giving the expected value of $Y_{\text{Sel}, n, X}(E/\mathbb{Q})$	Hyp. 4.1
$\theta_n(X, N)$	Moving ratio defined by $\pi_{\mathcal{S}_n}((X, X + N])/\pi_{\mathcal{E}}((X, X + N])$	Cor. 4.4
$Y_{\text{Hasse}, n, X}(s_E)$	Random variable with value 1 if $s_E \equiv 0 \in \text{III}(E/\mathbb{Q})$ , and 0 otherwise	Hyp. 5.3
$\rho_n(X)$	The function giving the expected value of $Y_{\text{Hasse}, n, X}(s_E)$	Hyp. 5.3
$\rho_n(X, N)$	Moving ratio approximating $\rho_n(X)$	Def. 5.17
$C_{s, t}^n(X)$	Covariance function of a certain products of random variables	Prop. 5.8
$\mathbb{E}_{s, t}^n(X)$	Expected value of a certain product of random variables	Rem. 5.12

TABLE 2. Notation defined and used throughout the paper.

In Table 2 we include a glossary of notation defined throughout the paper, together with a reference, for the reader's convenience. We also recall here a few definitions of probability concepts for the convenience of the reader. We say that a random variable  $Y$  follows a Bernoulli distribution  $B(1, p)$ , or  $Y \sim B(1, p)$ , if  $Y$  takes the value 1 with success probability of  $p$  and the value 0 with probability  $1 - p$ . The binomial distribution  $B(n, p)$  is the discrete probability distribution of the number of successes in a sequence of  $n$  independent yes/no experiments, each of which yields success with probability  $p$ . The expected value and variance of a discrete random variable  $Y$  that takes values  $y_1, \dots, y_k$  with probability  $p_1, \dots, p_k$  are defined respectively by

$$\mathbb{E}(Y) = \sum_{i=1}^k y_i \cdot p_i, \quad \text{Var}(Y) = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2.$$

The covariance of two random variables  $V, W$  is given by

$$\text{Cov}(V, W) = \mathbb{E}(VW) - \mathbb{E}(V) \cdot \mathbb{E}(W).$$

If  $\text{Cov}(V, W) = 0$ , then we say that  $V$  and  $W$  are uncorrelated random variables. If  $V$  and  $W$  are independent random variables, then  $\mathbb{E}(VW) = \mathbb{E}(V)\mathbb{E}(W)$  and, in particular,  $\text{Cov}(V, W) = 0$ . Also, we note here that if  $a$  and  $b$  are constants, then

$$\text{Var}(aV + bW) = a^2 \text{Var}(V) + b^2 \text{Var}(W) + 2ab \text{Cov}(V, W).$$

Finally, the standard error of the mean of random variables  $Y_1, \dots, Y_m$  is an estimator for the accuracy of the approximation of  $\frac{1}{m} \sum Y_i$  by  $\frac{1}{m} \sum \mathbb{E}(Y_i)$ , and it is defined as the square root of the variance of the mean of the variables. In other words, the standard error is given by

$$\sqrt{\text{Var}\left(\frac{1}{m} \sum_{i=1}^m Y_i\right)}.$$

If  $Y_1, \dots, Y_m$  are  $m$  independent random variables following the same distribution with mean  $\mu$  and standard deviation  $\sigma$ , then  $\text{SEM}(Y_1, \dots, Y_m) = (\frac{1}{m^2} \sum \text{Var}(Y_i))^{1/2} = (\text{Var}(Y_1)/m)^{1/2} = \sigma/\sqrt{m}$ .

### 3. THE NUMBER OF ELLIPTIC CURVES WITH (NAIVE) HEIGHT $\leq X$

Let  $E/\mathbb{Q}$  be an elliptic curve. We shall write each elliptic curve in a short Weierstrass model of the form  $y^2 = x^3 + Ax + B$  with  $A, B \in \mathbb{Z}$  and  $0 \neq 4A^3 + 27B^2$  such that  $\Delta_E$  is minimal in absolute value (minimal among all short Weierstrass models isomorphic to  $E$  over  $\mathbb{Q}$ ). In other words, we will be working with the set of elliptic curves

$$\mathcal{E} = \{E_{A,B} : y^2 = x^3 + Ax + B \mid A, B \in \mathbb{Z}, 4A^3 + 27B^2 \neq 0, \text{ and if } d^4 \mid A, d^6 \mid B, \text{ then } d = \pm 1\}.$$

Then, the (naive) height of  $E = E_{A,B} \in \mathcal{E}$  is defined by

$$\text{ht}(E_{A,B}) = \max\{4|A|^3, 27B^2\},$$

as used in [1], [4], and [21]. The BHKSSW database ([1]) contains data for all 238,764,310 elliptic curves up to height 26,998,673,868  $\approx 2.7 \cdot 10^{10}$ . While working on this project, we have gathered data for the curves  $y^2 = x^3 + Ax$ , for all fourth-power-free integers  $A \in [1, 10^6]$ , that is, about a million curves with  $j = 1728$ , up to height  $4 \cdot 10^{18}$ .

For each positive real number  $X$ , we define  $\mathcal{E}(X) = \{E \in \mathcal{E} : \text{ht}(E) \leq X\}$ , and  $\pi_{\mathcal{E}}(X) = \#\mathcal{E}(X)$ . Similarly, if  $0 \leq X_1 \leq X_2$ , we shall write  $\mathcal{E}([X_1, X_2])$  for the set  $\{E \in \mathcal{E} : X_1 \leq \text{ht}(E) \leq X_2\}$  and  $\pi_{\mathcal{E}}([X_1, X_2]) = \#\mathcal{E}([X_1, X_2])$  for its size (in particular,  $\mathcal{E}^X = \mathcal{E}([X, X])$  denotes the elliptic curves

of height *exactly*  $X$ , a set that can be empty depending on the value of  $X$ !). We shall refine an argument of Brumer ([4]) to estimate the value of  $\pi_{\mathcal{E}}(X)$ .

**Theorem 3.1.** *The number of elliptic curves of height up to  $X$  satisfies*

$$\left| \pi_{\mathcal{E}}(X) - \frac{2^{4/3} X^{5/6}}{3^{3/2} \zeta(10)} \right| \leq \frac{2X^{1/2}}{3^{3/2} \zeta(6)} + \frac{2^{4/3} X^{1/3}}{\zeta(4)} + \frac{2^{2/3} X^{1/6}}{\sqrt{3}} + O(X^{1/12+\varepsilon}),$$

for any  $\varepsilon > 0$ . In particular,  $\pi_{\mathcal{E}}(X) = \kappa X^{5/6} + O(X^{1/2})$  where the constant  $\kappa = 2^{4/3} \cdot (\zeta(10) \cdot 3^{3/2})^{-1} \approx 0.484462004349$ .

*Proof.* We follow Brumer's proof of Lemma 4.3 in [4]. In particular, we note that  $\mathcal{E}(X)$  is the set  $\mathcal{C}(X)$  of Brumer whose size we want to estimate. We modify Brumer's definition slightly to adjust for the change in height function (his height is given by  $\text{ht}(E_{A,B}) = \max\{|A|^3, B^2\}$ ):

$$\begin{aligned} \mathcal{D}(X) &= \{E_{A,B} : |A| \leq (X/4)^{1/3}, |B| \leq (X/27)^{1/2}\}, \quad \mathcal{D}'(X) = \mathcal{D}(X) - \{E_{0,0}\} \\ \mathcal{M}(X) &= \{E_{A,B} \in \mathcal{D}(X) : \text{if } d^4|A, d^6|B \text{ then } d = \pm 1\} \\ \mathcal{E}(X) &= \mathcal{C}(X) = \{E_{A,B} \in \mathcal{M}(X) : \Delta_{A,B} \neq 0\} \\ \mathcal{S}(X) &= \{E_{A,B} \in \mathcal{M}(X) : \Delta_{A,B} = 0\} \end{aligned}$$

so that  $\mathcal{M}(X) = \mathcal{E}(X) \sqcup \mathcal{S}(X)$ . Moreover, if we define an action  $d * E_{A,B} = E_{d^4 A, d^6 B}$ , then

$$\mathcal{D}'(X) = \bigsqcup_{d=1}^{\lfloor X^{1/12} \rfloor} d * \mathcal{M}(d^{-12} X).$$

Since  $\#d * \mathcal{M}(X) = \#\mathcal{M}(X)$ , it follows that  $\#\mathcal{D}'(X) = \sum_{d=1}^{\lfloor X^{1/12} \rfloor} \#\mathcal{M}(d^{-12} X)$ . We can use Möbius inversion to obtain:

$$\#\mathcal{E}(X) + \#\mathcal{S}(X) = \#\mathcal{M}(X) = \sum_{d=1}^{\lfloor X^{1/12} \rfloor} \mu(d) \cdot \#\mathcal{D}'(d^{-12} X)$$

Next, we note that

$$\mathcal{D}'(X) = \left( \left( 2 \left\lfloor (X/4)^{1/3} \right\rfloor + 1 \right) \left( 2 \left\lfloor (X/27)^{1/2} \right\rfloor + 1 \right) - 1 \right)$$

and

$$\begin{aligned} \#\mathcal{S}(X) &= \#\{(A, B) : A = -3C^2, B = 2C^3, |A| \leq (X/4)^{1/3}, \text{ with square-free } C\} \\ &\leq \{C : |C| \leq (X/4)^{1/6}/\sqrt{3}\} = \frac{2(X/4)^{1/6}}{\sqrt{3}} + 1. \end{aligned}$$

Therefore, putting everything together we obtain the following bound for  $|\#\mathcal{E}(X)|$ :

$$\begin{aligned} &\leq \#\mathcal{M}(X) + \#\mathcal{S}(X) \\ &= \sum_{d=1}^{\lfloor X^{1/12} \rfloor} \mu(d) \cdot \#\mathcal{D}'(d^{-12} X) + \frac{2(X/4)^{1/6}}{\sqrt{3}} + 1 \\ &= \sum_{d=1}^{\lfloor X^{1/12} \rfloor} \mu(d) \cdot \left( \left( 2 \left\lfloor \frac{(X/4)^{1/3}}{d^4} \right\rfloor + 1 \right) \left( 2 \left\lfloor \frac{(X/27)^{1/2}}{d^6} \right\rfloor + 1 \right) - 1 \right) + \frac{2^{2/3} X^{1/6}}{\sqrt{3}} + 1. \end{aligned}$$

Let us expand the sum as  $\sum_d \mu(d) \cdot (4\lfloor \alpha_d(X) \rfloor \lfloor \beta_d(X) \rfloor + 2\lfloor \alpha_d(X) \rfloor + 2\lfloor \beta_d(X) \rfloor)$ , with

$$\alpha_d(X) = \frac{(X/27)^{1/2}}{d^6}, \quad \beta_d(X) = \frac{(X/4)^{1/3}}{d^4}.$$

Now, note that  $\lfloor Y \rfloor \lfloor Z \rfloor = YZ - \{Y\}Z - \{Z\}Y + \{Y\}\{Z\}$ , where  $\{Y\} = Y - \lfloor Y \rfloor$  is the fractional part of  $Y$ . Thus,

$$\lfloor \alpha_d(X) \rfloor \lfloor \beta_d(X) \rfloor = \frac{X^{5/6}}{2^{2/3} 3^{3/2} d^{10}} - \{\alpha_d(X)\} \beta_d(X) - \{\beta_d(X)\} \alpha_d(X) + \{\alpha_d(X)\} \{\beta_d(X)\}.$$

Thus,

$$\begin{aligned} & 4\lfloor \alpha_d(X) \rfloor \lfloor \beta_d(X) \rfloor + 2\lfloor \alpha_d(X) \rfloor + 2\lfloor \beta_d(X) \rfloor \\ &= 4\lfloor \alpha_d(X) \rfloor \lfloor \beta_d(X) \rfloor + 2(\alpha_d(X) - \{\alpha_d(X)\}) + 2(\beta_d(X) - \{\beta_d(X)\}) \\ &= \frac{2^{4/3} X^{5/6}}{3^{3/2} d^{10}} + 2(1 - 2\{\beta_d(X)\})\alpha_d(X) + 2(1 - 2\{\alpha_d(X)\})\beta_d(X) \\ &\quad + 4\{\alpha_d(X)\}\{\beta_d(X)\} - 2\{\alpha_d(X)\} - 2\{\beta_d(X)\} \end{aligned}$$

Since  $0 \leq \{Y\}, |\mu(d)| \leq 1$ , we have  $|\sum_{d=1}^Z \mu(d)\{Y\}| \leq Z$ . Also, we know  $\sum_{d=1}^{\infty} \mu(d)/d^s = 1/\zeta(s)$ . It follows that

$$\left| \sum_{d=1}^{\lfloor X^{1/12} \rfloor} \mu(d)/d^s - 1/\zeta(s) \right| = \left| \sum_{d=\lfloor X^{1/12} \rfloor}^{\infty} \mu(d)/d^s \right| \leq \left| \sum_{d=\lfloor X^{1/12} \rfloor}^{\infty} 1/d^s \right| = O\left((X^{-1/12})^{s-1-\varepsilon}\right)$$

for any  $\varepsilon > 0$ . We shall write  $\zeta_{\text{tail}}(X, s) = \sum_{d=\lfloor X^{1/12} \rfloor}^{\infty} 1/d^s$ . In particular, we obtain

$$\begin{aligned} & \left| \sum_{d=1}^{\lfloor X^{1/12} \rfloor} \mu(d) \cdot 2(1 - 2\{\beta_d(X)\})\alpha_d(X) \right| = \left| 2 \sum_{d=1}^{\lfloor X^{1/12} \rfloor} \mu(d) \cdot \alpha_d(X) - 4 \sum_{d=1}^{\lfloor X^{1/12} \rfloor} \mu(d) \cdot \{\beta_d(X)\} \alpha_d(X) \right| \\ & \leq 2 \left| \sum_{d=1}^{\lfloor X^{1/12} \rfloor} \mu(d) \cdot \alpha_d(X) \right| = 2 \cdot \left( \frac{1}{\zeta(6)} + \zeta_{\text{tail}}(X, 6) \right) (X/27)^{1/2} \\ & = \frac{2(X/27)^{1/2}}{\zeta(6)} + \zeta_{\text{tail}}(X, 6) \cdot (X/27)^{1/2} = \frac{2(X/27)^{1/2}}{\zeta(6)} + O(X^{1/12+\varepsilon}), \end{aligned}$$

and similarly

$$\left| \sum_{d=1}^{\lfloor X^{1/12} \rfloor} \mu(d) \cdot 2(1 - 2\{\alpha_d(X)\})\beta_d(X) \right| \leq \frac{2(X/4)^{1/3}}{\zeta(4)} + \zeta_{\text{tail}}(X, 4) \cdot (X/4)^{1/3} = \frac{2(X/4)^{1/3}}{\zeta(4)} + O(X^{1/12+\varepsilon}).$$

Hence,

$$\begin{aligned} \left| \#\mathcal{E}(X) - \frac{2^{4/3} X^{5/6}}{3^{3/2} \zeta(10)} \right| & \leq \frac{2X^{1/2}}{3^{3/2} \zeta(6)} + \frac{2^{4/3} X^{1/3}}{\zeta(4)} + \frac{2^{2/3} X^{1/6}}{\sqrt{3}} \\ & \quad + \zeta_{\text{tail}}(X, 6) \cdot (X/27)^{1/2} + \zeta_{\text{tail}}(X, 4) \cdot (X/4)^{1/3} + 8 \cdot X^{1/12} \\ & = \frac{2X^{1/2}}{3^{3/2} \zeta(6)} + \frac{2^{4/3} X^{1/3}}{\zeta(4)} + \frac{2^{2/3} X^{1/6}}{\sqrt{3}} + O(X^{1/12+\varepsilon}), \end{aligned}$$



as desired.  $\square$

**Remark 3.2.** Using the BHKSSW database, we have calculated the values of  $\pi_{\mathcal{E}}(X)$  up to  $2.7 \cdot 10^{10}$  in  $0.25 \cdot 10^9$  intervals. We have found (using SageMath, [26]) the best-fit model of the form  $C \cdot X^{5/6}$  for these data points, and found that the best constant is  $C \approx 0.48447036$  in agreement with Brumer's constant ( $C$  and  $\kappa$  differ by  $8.35 \cdot 10^{-6}$ ).

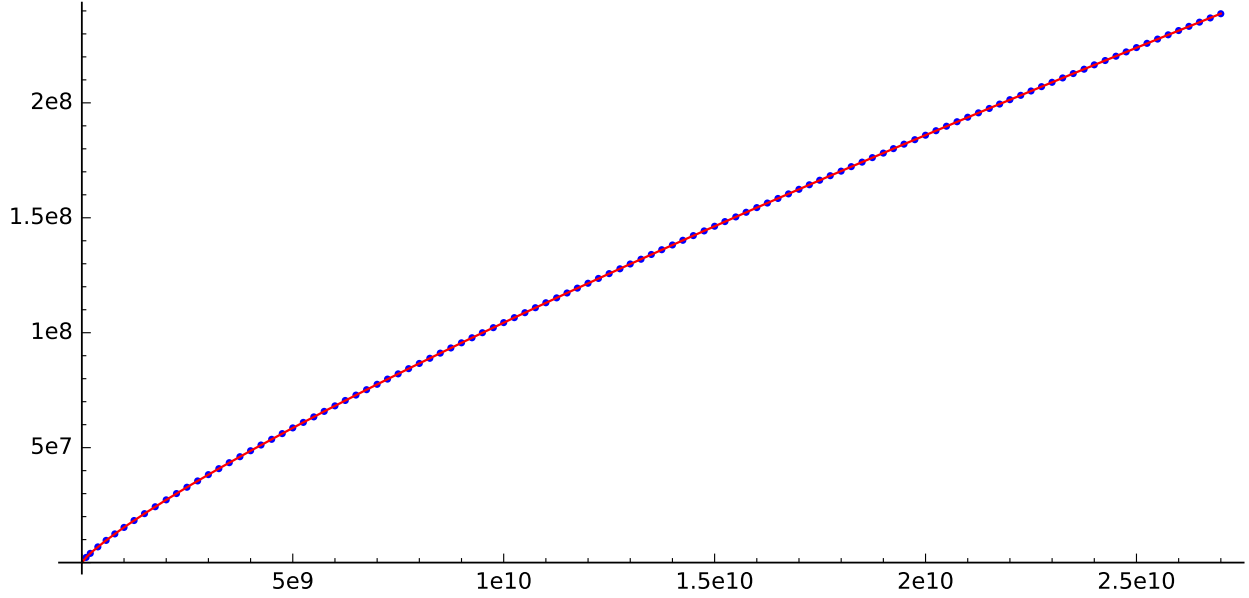


FIGURE 1. Values of  $\pi_{\mathcal{E}}(X)$  from the BHKSSW database (blue dots), and the function  $0.48447036 \cdot X^{5/6}$  (in red).

**Remark 3.3.** According to Theorem 3.1, the number of curves in the height interval  $(X, X + N]$  is, approximately,

$$\begin{aligned} \pi_{\mathcal{E}}((X, X + N]) &= \pi_{\mathcal{E}}(X + N) - \pi_{\mathcal{E}}(X) \\ &\approx \kappa \cdot ((X + N)^{5/6} - X^{5/6}) \\ &= \frac{5\kappa}{6} \int_X^{X+N} \frac{1}{H^{1/6}} dH \approx \frac{5\kappa}{6} \cdot \frac{N}{X^{1/6}}, \end{aligned}$$

where  $5\kappa/6 \approx 0.403718336957$ , and the last approximation is valid for large  $X$  such that  $X \gg N \geq 0$ . However, the error in this approximation is still of the order  $O(X^{1/2})$ , so the error can be quite large. For instance,

$$\begin{aligned} \pi([2 \cdot 10^{10}, 2 \cdot 10^{10} + 0.25 \cdot 10^9]) &= 1,955,593 \approx 1,937,225.394 \dots = \frac{5\kappa}{6} \cdot \frac{0.25 \cdot 10^9 + 1}{(2 \cdot 10^{10})^{1/6}} \\ \pi([2.5 \cdot 10^{10}, 2.5 \cdot 10^{10} + 0.25 \cdot 10^9]) &= 1,852,352 \approx 1,866,502.107 \dots = \frac{5\kappa}{6} \cdot \frac{0.25 \cdot 10^9 + 1}{(2.5 \cdot 10^{10})^{1/6}} \end{aligned}$$

Nonetheless, we shall prove below (Corollary 3.4) that the approximation  $\pi_{\mathcal{E}}((X, X+N]) \approx \frac{5\kappa}{6} \cdot \frac{N}{X^{1/6}}$  works *on average* with error going to zero as  $X$  goes to infinity (as long as  $X > N^2$ ).

We also point out here that if we want  $\pi_{\mathcal{E}}((X, X+N])$  to be approximately constant as  $X \rightarrow \infty$ , then we need  $N = N(X) \asymp C \cdot X^{1/6}$ . For instance, if we want  $\pi_{\mathcal{E}}((X, X+N(X)]) \approx 10^t$ , then we should have  $N = N(X) = (6 \cdot 10^t / 5\kappa) \cdot X^{1/6}$ , where  $6/5\kappa \approx 2.476974436029$ .

**Corollary 3.4.** *Let  $N \geq 1$  be fixed, and suppose  $X > N^2$ . Then, on average,*

$$\pi_{\mathcal{E}}((X, X+N]) \approx \int_X^{X+N} \frac{5\kappa/6}{H^{1/6}} dH + O\left(\frac{N}{X^{1/2}}\right) \approx \frac{5\kappa}{6} \cdot \frac{N}{(X+N)^{1/6}} + O\left(\frac{N}{X^{1/2}}\right),$$

*in the sense that if  $X_i = i \cdot N$  for  $i = 0, \dots, \lfloor X/N \rfloor$ , then*

$$\frac{1}{\lfloor X/N \rfloor} \cdot \sum_{i=0}^{\lfloor X/N \rfloor - 1} \left( \pi_{\mathcal{E}}((X_i, X_{i+1}]) - \int_{X_i}^{X_{i+1}} \frac{5\kappa/6}{H^{1/6}} dH \right) = O\left(\frac{N}{X^{1/2}}\right)$$

*and*

$$\left| \frac{1}{\lfloor X/N \rfloor} \cdot \sum_{i=0}^{\lfloor X/N \rfloor - 1} \left( \pi_{\mathcal{E}}((X_i, X_{i+1}]) - \frac{5\kappa}{6} \cdot \frac{N}{X_{i+1}^{1/6}} \right) \right| = O\left(\frac{N}{X^{1/2}}\right).$$

*In particular,  $\pi_{\mathcal{E}}^X = \pi_{\mathcal{E}}([X, X]) = \pi_{\mathcal{E}}((X-1, X]) \approx (5\kappa/6)/X^{1/6} + O(X^{-1/2})$ , on average.*

*Proof.* Let  $N \geq 1$  be fixed, let  $X > N^2$  be fixed, and let us define  $X_i = i \cdot N$  for  $i = 0, \dots, \lfloor X/N \rfloor$ . Then

$$\begin{aligned} & \frac{1}{\lfloor X/N \rfloor} \cdot \sum_{i=0}^{\lfloor X/N \rfloor - 1} \left( \pi_{\mathcal{E}}((X_i, X_{i+1}]) - \int_{X_i}^{X_{i+1}} \frac{5\kappa/6}{H^{1/6}} dH \right) \\ &= \frac{1}{\lfloor X/N \rfloor} \left( \pi_{\mathcal{E}}([1, \lfloor X/N \rfloor N]) - \int_0^{\lfloor X/N \rfloor N} \frac{5\kappa/6}{H^{1/6}} dH \right) \\ &= \frac{1}{\lfloor X/N \rfloor} \left( \pi_{\mathcal{E}}(\lfloor X/N \rfloor N) - \kappa \cdot (\lfloor X/N \rfloor N)^{5/6} \right) \\ &= \frac{1}{\lfloor X/N \rfloor} \cdot O((\lfloor X/N \rfloor N)^{1/2}) = O\left(\frac{X^{1/2}}{X/N}\right) = O\left(\frac{N}{X^{1/2}}\right), \end{aligned}$$

by Theorem 3.1. Now it follows that

$$\begin{aligned} & \left| \frac{1}{\lfloor X/N \rfloor} \cdot \sum_{i=0}^{\lfloor X/N \rfloor - 1} \left( \pi_{\mathcal{E}}((X_i, X_{i+1}]) - \frac{5\kappa}{6} \cdot \frac{N}{X_{i+1}^{1/6}} \right) \right| \\ &= \left| \frac{1}{\lfloor X/N \rfloor} \cdot \sum_{i=0}^{\lfloor X/N \rfloor - 1} \left( \pi_{\mathcal{E}}((X_i, X_{i+1}]) - \int_{X_i}^{X_{i+1}} \frac{5\kappa/6}{H^{1/6}} dH + \int_{X_i}^{X_{i+1}} \frac{5\kappa/6}{H^{1/6}} dH - \frac{5\kappa}{6} \cdot \frac{N}{X_{i+1}^{1/6}} \right) \right| \end{aligned}$$

$$\begin{aligned}
&\leq O\left(\frac{N}{X^{1/2}}\right) + \left| \frac{1}{\lfloor X/N \rfloor} \cdot \sum_{i=0}^{\lfloor X/N \rfloor - 1} \left( \int_{X_i}^{X_{i+1}} \frac{5\kappa/6}{H^{1/6}} dH - \frac{5\kappa}{6} \cdot \frac{N}{X_{i+1}^{1/6}} \right) \right| \\
&= O\left(\frac{N}{X^{1/2}}\right) + \left| \frac{1}{\lfloor X/N \rfloor} \cdot (5\kappa/6)N \cdot \sum_{i=0}^{\lfloor X/N \rfloor - 1} \left( \frac{1}{X_i^{1/6}} - \frac{1}{X_{i+1}^{1/6}} \right) \right| \\
&= O\left(\frac{N}{X^{1/2}}\right) + \left| \frac{1}{\lfloor X/N \rfloor} \cdot (5\kappa/6)N \cdot \left( 1 - \frac{1}{(\lfloor X/N \rfloor N)^{1/6}} \right) \right| \\
&= O\left(\frac{N}{X^{1/2}}\right) + O\left(\frac{N^2}{X}\right) = O\left(\frac{N}{X^{1/2}}\right),
\end{aligned}$$

where we have used the fact that  $X > N^2$ , and therefore  $1 \geq (N^2/X)^{1/2} > N^2/X > 0$ , and if  $f(x)$  is decreasing, then

$$0 \leq \int_a^b f(x)dx - (b-a)f(b) \leq (b-a)(f(a) - f(b)),$$

with  $b-a = X_{i+1} - X_i = N$  and  $f(x) = (5\kappa/6)/x^{1/6}$ .  $\square$

**Remark 3.5.** For each  $n \geq r \geq 0$  and  $X \geq N \geq 0$ , we would like to be able to give estimates of the quantities  $\pi_{\mathcal{E}}((X, X+N])$ ,  $\pi_{\mathcal{S}_n}((X, X+N])$  and  $\pi_{\mathcal{R}_r}((X, X+N])$ . However, Corollary 3.4 says that this is not possible, and the best we hope for are results *on average*. From now on, if  $f = f(X, N)$  and  $g = g(X, N)$  are functions, then we say that  $f \approx g$  on average, for  $X \gg N$ , if

$$(1) \quad \frac{1}{\lfloor X/N \rfloor} \cdot \left| \sum_{i=0}^{\lfloor X/N \rfloor - 1} f(X_i, N) - g(X_i, N) \right|$$

goes to zero as  $X \rightarrow \infty$ , where  $X_i = i \cdot N$  for  $i = 0, \dots, \lfloor X/N \rfloor$ . The condition  $X \gg N$  is to be specified every time, but in general we will assume  $X > N^2$  as in Corollary 3.4. If we want to be more specific about the error terms, we shall say  $f \approx g + O(h(X, N))$  if the quantity in Eq. (1) is  $O(h(X, N))$ .

We finish this section by stating an immediate consequence of the definition of  $f \approx g$  on average.

**Lemma 3.6.** *Let  $f = f(X, N)$  and  $g = g(X, N)$  be functions such that  $f \approx g$  on average for  $X \gg N \geq 0$ , and let  $a(X)$  be a function such that  $|a(X)| \leq 1$  for all  $X \geq x_0$ , for some  $x_0 \in \mathbb{R}$ . Then,  $a(X) \cdot f(X, N) \approx a(X) \cdot g(X, N)$  on average. More generally, suppose  $f \approx g + O(h)$  for some other functions  $h = h(X, N)$ .*

- (1) *If  $a(X) = O(b(X))$  for some other functions  $h = h(X, N)$  and  $b = b(X)$ , then  $a \cdot f \approx a \cdot g + b \cdot O(h) \approx a \cdot g + O(b \cdot h)$  on average.*
- (2) *Suppose  $g(X, N) = \int_X^{X+N} \hat{g}(H) dH$  for some integrable function  $\hat{g}$ , and suppose that  $a(X)$  is also integrable with  $a = O(b(X))$ . Then,*

$$a(X) \cdot f(X, N) \approx \int_X^{X+N} a(H) \cdot \hat{g}(H) dH + b(X) \cdot O(h(X)) \approx \int_X^{X+N} a(H) \cdot \hat{g}(H) dH + O(b \cdot h)$$

*on average.*

4. THE NUMBER OF CURVES WITH SELMER RANK  $n$  UP TO HEIGHT  $X$ 

Let  $\text{Sel}_2(E/\mathbb{Q})$  be the 2-Selmer group of  $E/\mathbb{Q}$  and let  $\text{III}(E/\mathbb{Q})[2]$  be the 2-torsion subgroup of the Tate-Shafarevich group of  $E/\mathbb{Q}$  (as defined in [24], Chapter X) which fit in a short exact sequence

$$(2) \quad 0 \longrightarrow E(\mathbb{Q})/2E(\mathbb{Q}) \xrightarrow{\delta_E} \text{Sel}_2(E/\mathbb{Q}) \longrightarrow \text{III}(E/\mathbb{Q})[2] \longrightarrow 0$$

As in [14], we shall refer to the quantity

$$\text{rank}_{\mathbb{Z}/2\mathbb{Z}}(\text{Sel}_2(E/\mathbb{Q})/(E(\mathbb{Q})_{\text{tors}}/2E(\mathbb{Q})_{\text{tors}})) = \text{rank}_{\mathbb{Z}/2\mathbb{Z}}(\text{Sel}_2(E/\mathbb{Q})) - \text{rank}_{\mathbb{Z}/2\mathbb{Z}}(E(\mathbb{Q})[2])$$

as the 2-Selmer rank (or, simply, Selmer rank) of  $E/\mathbb{Q}$ , and will denote it by  $\text{selrank}(E(\mathbb{Q}))$ . We note here that the exact sequence above implies that  $\text{rank}(E(\mathbb{Q})) \leq \text{selrank}(E(\mathbb{Q}))$  for all elliptic curves. We define

$$\mathcal{S}_n = \{E \in \mathcal{E} : \text{selrank}(E(\mathbb{Q})) = n\},$$

and we will denote by  $\mathcal{S}_n(X)$  those curves in  $\mathcal{S}_n$  of height up to  $X$ , and  $\pi_{\mathcal{S}_n}(X) = \#\mathcal{S}_n(X)$ . Imitating the notation in the previous section, we shall also write  $\mathcal{S}_n([X_1, X_2])$  and  $\pi_{\mathcal{S}_n}([X_1, X_2])$  when referring to curves in  $\mathcal{S}_n$  in the height interval  $[X_1, X_2]$ , and will abbreviate  $\mathcal{S}_n^X = \pi_{\mathcal{S}_n}([X, X])$ . Poonen and Rains ([22]) have conjectured a value for the limit  $s_n = \lim_{X \rightarrow \infty} \pi_{\mathcal{S}_n}(X)/\pi_{\mathcal{E}}(X)$ , namely

$$s_n = \text{Prob}(\text{selrank}(E(\mathbb{Q})) = n) = \left( \prod_{j \geq 0} \frac{1}{1 + 2^{-j}} \right) \cdot \left( \prod_{k=1}^n \frac{2}{2^k - 1} \right),$$

and, in fact, they conjecture a similar distribution for  $p$ -Selmer groups of rank  $n$ , and any prime  $p$ . This probability has been shown to hold for quadratic twists of certain elliptic curves (see [14], [15], [25], and [16]). The value of the constant  $s_0 = \prod_{j \geq 0} (1 + 2^{-j})^{-1}$  is approximately 0.20971122, and we have included approximations of  $s_n$  for  $n = 1, \dots, 6$  for future reference in Table 1.

$s_0$	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	$s_6$
0.20971122	0.41942244	0.27961496	0.07988998	0.01065199	0.00068722	0.00002181

TABLE 3. Values of  $s_n = \text{Prob}(\text{selrank}(E(\mathbb{Q})) = n)$

For our purposes, we are interested in the behavior of the function  $\mathcal{S}_n(X)$ , but we are even more interested in the conditional probability

$$\text{pSel}_n(X) = \text{Prob}(\text{selrank}(E(\mathbb{Q})) = n \mid \text{ht}(E) = X) = \#\mathcal{S}_n^X / \#\mathcal{E}^X,$$

when  $\#\mathcal{E}^X \neq 0$ . In other words, we would like to know the probability that a curve  $E$  of height  $X$  has Selmer rank  $n$ , and in order to study  $\text{pSel}_n(X)$  we introduce the following probabilistic hypothesis to be tested.

**Hypothesis 4.1** (Hypothesis A, or  $H_A$ ). *Let  $n \geq 0$ , let  $X \geq 0$ , and let  $Y_{\text{Sel},n,X} : \mathcal{E}^X \rightarrow \{0, 1\}$  be the function that takes values*

$$Y_{\text{Sel},n,X}(E/\mathbb{Q}) = \begin{cases} 1 & \text{if } \text{selrank}(E(\mathbb{Q})) = n, \\ 0 & \text{otherwise.} \end{cases}$$

There exists a function  $\theta_n(X)$ , continuous for  $X > 0$ , such that if  $\mathcal{E}^X \neq \emptyset$ , then  $Y_{\text{Sel},n,X}(E/\mathbb{Q})$  behaves as a random variable that follows a Bernoulli distribution with probability  $\theta_n(X)$ , such that  $\lim_{X \rightarrow \infty} \theta_n(X) = s_n$ .

Moreover, if  $E/\mathbb{Q}$  and  $E'/\mathbb{Q}$  are non-isogenous, then the random variables  $Y_{\text{Sel},n,X}(E/\mathbb{Q})$  and  $Y_{\text{Sel},n,X}(E'/\mathbb{Q})$  are independent (and, therefore, uncorrelated).

In other words, Hypothesis A claims that  $Y_{\text{Sel},n,X}(E/\mathbb{Q}) \sim B(1, \theta_n(X))$ , i.e.,  $Y_{\text{Sel},n,X}$  follows a binomial distribution with one trial with probability  $\theta_n(X)$ , where  $E/\mathbb{Q}$  is an elliptic curve of Selmer rank  $n$  and height  $X$ .

**Remark 4.2.** When considering large sample sets, we can replace “non-isogenous” in  $H_A$  by “non-isomorphic”, because by a theorem of Kenku ([17]), an elliptic curve  $E/\mathbb{Q}$  is isogenous to at most 8 non-isomorphic elliptic curves over  $\mathbb{Q}$ . Below, we will simply neglect the (small) error introduced by possibly considering isogenous curves in our sample sets.

**Corollary 4.3.** Assume  $H_A$ , and let  $\mathfrak{E} = \{E_1, \dots, E_m\}$  be a set of distinct elliptic curves in  $\mathcal{E}^X$  chosen at random. Then, the number of curves in  $\mathfrak{E}$  of Selmer rank  $n$  follows a binomial distribution  $B(m, \theta_n(X))$ . In particular the expected value of  $\#(\mathfrak{E} \cap \mathcal{S}_n)/\#\mathfrak{E}$  is  $\theta_n(X)$  with standard error  $\sqrt{\frac{1}{m} \theta_n(X)(1 - \theta_n(X))}$ . More generally, if  $\{E_1, \dots, E_m\}$  are distinct elliptic curves in  $\mathcal{E}$ , with  $E_i$  of height  $X_i$  for  $i = 1, \dots, m$ , and chosen at random, then

$$\mathbb{E}(\#(\mathfrak{E} \cap \mathcal{S}_n)/\#\mathfrak{E}) = \frac{1}{m} \sum_{i=1}^m \theta_n(X_i)$$

with standard error  $\sqrt{\frac{1}{m^2} \sum \theta_n(X_i)(1 - \theta_n(X_i))}$ .

*Proof.* Let us assume  $H_A$  and let us first show the most general case. Let  $E_1, \dots, E_m$  be distinct elliptic curves in  $\mathcal{E}$  of height  $X_1, \dots, X_m$ , respectively, chosen at random. In particular, by  $H_A$ , each random variable  $Y_{\text{Sel},n,X_i}(E_i/\mathbb{Q}) \sim B(1, \theta_n(X_i))$ , and since the curves are distinct,  $H_A$  says that the random variables are independent (see Remark 4.2). Then, the number of elements in  $\#\mathfrak{E} \cap \mathcal{S}_n$  can be expressed as

$$t = \#\mathfrak{E} \cap \mathcal{S}_n = \sum_{i=1}^m Y_{\text{Sel},n,X_i}(E_i/\mathbb{Q}).$$

It follows that the expected value of  $t$  is

$$\mathbb{E}(t) = \mathbb{E}\left(\sum_{i=1}^m Y_{\text{Sel},n,X_i}(E_i/\mathbb{Q})\right) = \sum_{i=1}^m \mathbb{E}(Y_{\text{Sel},n,X_i}(E_i/\mathbb{Q})) = \sum_{i=1}^m \theta_n(X_i),$$

and so the expected value of  $\#\mathfrak{E} \cap \mathcal{S}_n/\#\mathfrak{E}$  is  $\frac{1}{m} \sum \theta_n(X_i)$ . The standard error of the approximation of  $t/m$  by  $\frac{1}{m} \sum \theta_n(X_i)$  is given by the square root of the variance of  $t/m$ . We compute

$$\text{Var}\left(\frac{1}{m} \sum_{i=1}^m Y_{\text{Sel},n,X_i}(E_i/\mathbb{Q})\right) = \frac{1}{m^2} \sum_{i=1}^m \text{Var}(Y_{\text{Sel},n,X_i}(E_i/\mathbb{Q})) = \frac{1}{m^2} \sum_{i=1}^m \theta_n(X_i)(1 - \theta_n(X_i)),$$

where we have used the fact that  $Y_{\text{Sel},n,X_i}(E_i/\mathbb{Q})$  are independent, which implies they are uncorrelated, and therefore the covariance terms vanish. Thus, the standard error is  $\sqrt{\frac{1}{m^2} \sum \theta_n(X_i)(1 - \theta_n(X_i))}$  as claimed.

Now, if  $X = X_1 = \dots = X_m$ , then  $t = \#\mathfrak{E} \cap \mathcal{S}_n = \sum_{i=1}^m Y_{\text{Sel},n,X}(E_i/\mathbb{Q})$  follows a binomial  $B(m, \theta_n(X))$ , with mean  $m \cdot \theta_n(X)$  and variance  $\frac{1}{m} \theta_n(X)(1 - \theta_n(X))$ , so the expected value of  $t/m$  is  $\theta_n(X)$  with standard error  $\sqrt{\frac{1}{m} \theta_n(X)(1 - \theta_n(X))}$ , as desired.  $\square$

**Corollary 4.4.** *If we assume  $H_A$ , then  $\sum_{n=0}^{\infty} \theta_n(X) = 1$ . Moreover, for  $N \geq 1$ , if we define*

$$\theta_n(X, N) = \frac{\pi_{\mathcal{S}_n}((X, X + N])}{\pi_{\mathcal{E}}((X, X + N])},$$

then

(1) *The expected value of  $\pi_{\mathcal{S}_n}((X, X + N])$  is given by the formula*

$$\mathbb{E}(\pi_{\mathcal{S}_n}((X, X + N])) \approx \frac{5\kappa}{6} \cdot \int_X^{X+N} \frac{\theta_n(H)}{H^{1/6}} dH + O\left(\frac{N}{X^{1/2}}\right),$$

*on average (see Remark 3.5).*

(2) *For  $X > N^2 \geq 0$ , we have that the expected value of  $\theta_n(X, N)$  is, on average,  $\theta_n(X) + O(X^{-1/3})$ , with a standard error given on average by*

$$\sqrt{\frac{6 \cdot (X + N)^{1/6} \cdot \theta_n(X)(1 - \theta_n(X))}{5\kappa N}} + O\left(\frac{1}{NX^{1/6}}\right).$$

(3) *Let  $N = N(X)$  be a function of  $X$ . Then,  $\lim_{X \rightarrow \infty} \theta_n(X, N(X)) = s_n$  as long as the growth condition  $\lim_{X \rightarrow \infty} X^{1/6}/N(X) = 0$  is satisfied.*

*Proof.* Let  $E/\mathbb{Q}$  be a fixed elliptic curve of height  $X$ . Since  $\text{selrank}(E(\mathbb{Q}))$  takes precisely one value (a non-negative number), it follows that

$$\sum_{n=0}^{\infty} \theta_n(X) = \sum_{n \geq 0} \text{Prob}(Y_{\text{Sel},n,X}(E/\mathbb{Q}) = 1) = 1,$$

by the laws of probability. For part (1) of the statement, we note that

$$\pi_{\mathcal{S}_n}((X, X + N]) = \sum_{E \in \mathcal{E}((X, X + N])} Y_{\text{Sel},n,X}(E/\mathbb{Q}).$$

and therefore we may use Corollary 4.3 to obtain the expected value.

$$\begin{aligned} \mathbb{E}(\pi_{\mathcal{S}_n}((X, X + N])) &= \sum_{E \in \mathcal{E}((X, X + N])} \theta_n(\text{ht}(E)) \\ &= \sum_{H=X+1}^{X+N} \sum_{E \in \mathcal{E}([H, H])} \theta_n(H) = \sum_{H=X+1}^{X+N} \pi_{\mathcal{E}}([H, H]) \cdot \theta_n(H) \end{aligned}$$

Corollary 3.4 says that  $\pi_{\mathcal{E}}([H, H]) \approx \frac{5\kappa}{6} \cdot \int_{H-1}^H \frac{1}{T^{1/6}} dT + O(H^{-1/2})$  on average (see Remark 3.5 for the definition of the term “on average” in this context). Since  $|\theta_n(X)| \leq 1$  for all  $X$ , Lemma 3.6

implies that  $\pi_{\mathcal{E}}([H, H])\theta_n(H) \approx \frac{5\kappa}{6} \cdot \int_{H-1}^H \frac{\theta_n(T)}{T^{1/6}} dT + O(H^{-1/2})$  on average. Thus,

$$\begin{aligned} \mathbb{E}(\pi_{\mathcal{S}_n}((X, X+N])) &\approx \sum_{H=X+1}^{X+N} \left( \frac{5\kappa}{6} \cdot \int_{H-1}^H \frac{\theta_n(T)}{T^{1/6}} dT + \theta_n(H) \cdot O(H^{-1/2}) \right) \\ &= \frac{5\kappa}{6} \cdot \int_X^{X+N} \frac{\theta_n(H)}{H^{1/6}} dH + \theta_n(X) \cdot O\left(\frac{N}{X^{1/2}}\right), \end{aligned}$$

on average. For part (2), we use instead the second part of Corollary 3.4 which says  $\pi_{\mathcal{E}}([H, H]) \approx \frac{(5\kappa/6)}{H^{1/6}} + O(H^{-1/2})$ , on average. Thus,  $\mathbb{E}(\pi_{\mathcal{S}_n}((X, X+N]))$  is, on average, given by

$$\approx \frac{5\kappa}{6} \cdot \frac{N}{(X+N)^{1/6}} \cdot \theta_n(X) + O\left(\frac{N}{X^{1/2}}\right) \approx \pi_{\mathcal{E}}((X, X+N]) \cdot \theta_n(X) + O\left(\frac{N}{X^{1/2}}\right).$$

Thus,  $\mathbb{E}(\theta_n(X, N)) \approx \theta_n(X) + O\left(\frac{(X+N)^{1/6}}{X^{1/2}}\right) = \theta_n(X) + O(X^{-1/3})$  as claimed, since we are assuming that  $X > N^2$ . Similarly, the variance is given by

$$\begin{aligned} \text{Var}(\theta_n((X, X+N])) &= \frac{1}{(\pi_{\mathcal{E}}((X, X+N]))^2} \sum_{E \in \mathcal{E}((X, X+N])} \theta_n(\text{ht}(E))(1 - \theta_n(\text{ht}(E))) \\ &= \frac{1}{(\pi_{\mathcal{E}}((X, X+N]))^2} \sum_{H=X+1}^{X+N} \sum_{E \in \mathcal{E}([H, H])} \theta_n(H)(1 - \theta_n(H)) \\ &\approx \frac{5\kappa/6}{(\pi_{\mathcal{E}}((X, X+N]))^2} \left( \int_X^{X+N} \frac{\theta_n(H)(1 - \theta_n(H))}{H^{1/6}} dH + O\left(\frac{N}{X^{1/2}}\right) \right), \end{aligned}$$

on average. If  $X > N^2 \geq 0$ , then we obtain

$$\begin{aligned} \text{Var}(\theta_n((X, X+N])) &\approx \frac{5\kappa/6}{(\pi_{\mathcal{E}}((X, X+N]))^2} \cdot \frac{N}{(X+N)^{1/6}} \cdot \theta_n(X)(1 - \theta_n(X)) \\ &\approx \frac{\theta_n(X)(1 - \theta_n(X))}{\pi_{\mathcal{E}}((X, X+N]))} \approx \frac{6}{5\kappa} \cdot \frac{(X+N)^{1/6}}{N} \cdot \theta_n(X)(1 - \theta_n(X)), \end{aligned}$$

on average, with error term

$$\left( \frac{N}{(X+N)^{1/6}} \right)^{-2} \cdot O\left(\frac{N}{X^{1/2}}\right) = O\left(\frac{(X+N)^{1/3}}{NX^{1/2}}\right) = O\left(\frac{1}{NX^{1/6}}\right).$$

Finally, the maximum value of the function  $f(x) = x(1-x)$  in  $[0, 1]$  is  $1/4$ , and therefore standard error can be bounded by  $\sqrt{3(X+N)^{1/6}/(10N\kappa)} + O(1/(NX^{1/6}))$ . If we choose  $N = N(X)$  that grows faster than  $X^{1/6}$  (but not faster than  $X^{1/2}$ , so that  $X > N^2$ ), then the standard error of the approximation  $\theta_n(X, N) \approx \theta_n(X)$  goes to zero as  $X \rightarrow \infty$ . Hence,

$$\lim_{X \rightarrow \infty} \theta_n(X, N(X)) = \lim_{X \rightarrow \infty} \theta_n(X) = s_n.$$

This concludes the proof of the corollary.  $\square$

In order to test Hypothesis A, we have used the BHKSSW data to estimate the function  $\theta_n(X)$  using the moving ratios  $\theta_n(X, N)$  of Corollary 4.4. We have plotted values of  $\theta_n(X, 0.025 \cdot 10^9)$  for  $n = 1, \dots, 5$  using the BHKSSW database, and the graphs can be found in Figure 2.

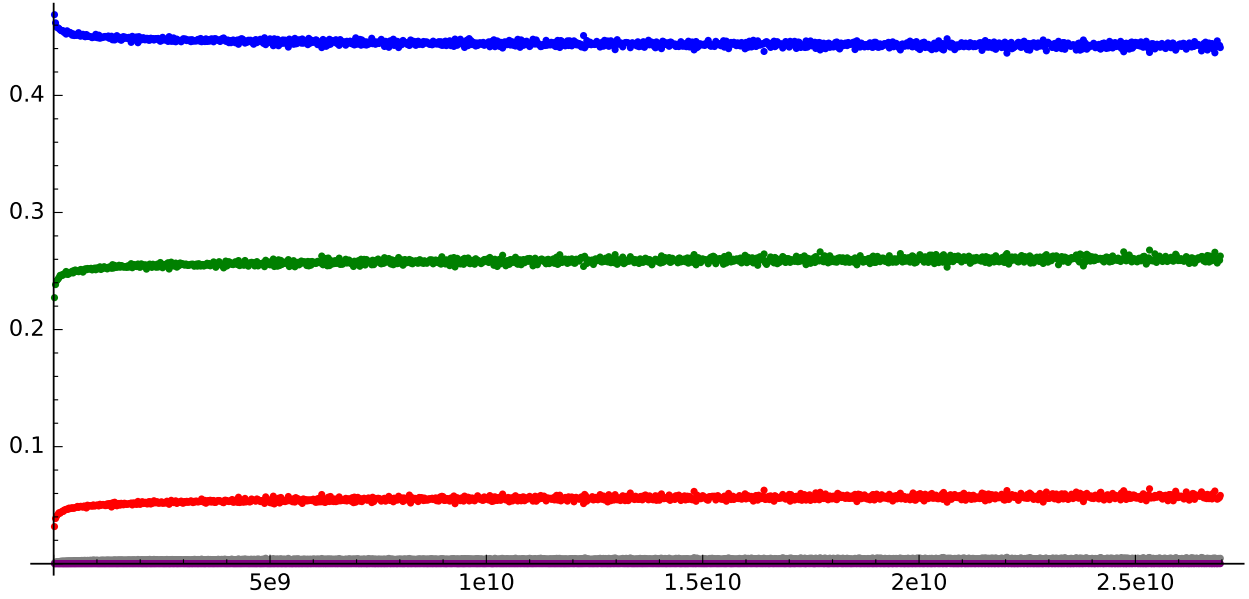


FIGURE 2. Graphs of the moving ratios  $\theta_n(X, 0.025 \cdot 10^9)$  for  $n = 1$  (blue), 2 (green), 3 (red), 4 (gray), 5 (purple).

In Table 4 we record the last values of  $\theta_n(X, 0.025 \cdot 10^9)$  that appear in the graphs (which correspond to  $X \approx 2.6975 \cdot 10^{10}$ ). We also record the values of  $\pi_{S_n}$  in  $[2.6975 \cdot 10^{10}, 2.7 \cdot 10^{10}]$ . The total number of elliptic curves in the same interval is 182,823.

$n$	1	2	3	4	5
$\pi_{S_n}([2.6975 \cdot 10^{10}, 2.7 \cdot 10^{10}])$	80,996	47,427	10,556	821	29
$\theta_n(2.6975 \cdot 10^{10}, 0.025 \cdot 10^9)$	0.44083621	0.26278969	0.05835152	0.00463836	0.00008751
$s_n$	0.41942244	0.27961496	0.07988998	0.01065199	0.00068722

TABLE 4. The number of curves of Selmer rank  $1 \leq n \leq 5$ , and the values of  $\theta_n(X, N)$  in the interval  $[2.6975 \cdot 10^{10}, 2.7 \cdot 10^{10}]$ , together with the values of  $s_n$ .

Finally, we have found (using SageMath) best-fit models for the data of  $\theta_n(X, N)$  of the form

$$\theta_n(X, N) \approx \frac{s_n}{1 + C_n X^{-e_n}}.$$

and we provide the values of  $C_n$  and  $e_n$  in Table 5. The

The models constructed above are remarkably good approximations of the values of  $\theta_n(X, 0.025 \cdot 10^9)$ , at least up to height  $2.7 \cdot 10^{10}$ . See Figures 3 and 4. We conjecture that if  $H_A$  is true, then  $\theta_n(X)$  are in fact well approximated by these models.



$n$	1	2	3	4	5
$C_n$	-0.40116957	1.41108621	11.18222736	179.71749981	95474.85098037
$e_n$	0.08540201	0.12348659	0.14061542	0.20339670	0.39937065

TABLE 5. The parameters of the best-fit models  $\theta_n(X, N) \approx s_n/(1 + C_n X^{-e_n})$ .

**Conjecture 4.5.** Hypothesis  $H_A$  holds and, for each  $n \geq 1$ , there are constants  $C_n$  and  $e_n$  such that  $\theta_n(X) \approx \frac{s_n}{1 + C_n X^{-e_n}}$ . Moreover, if  $n = 1, \dots, 5$ , then the approximate values of  $C_n$  and  $e_n$  are as in Table 5.

**Remark 4.6.** Let us assume Conjecture 4.5, and let us use Corollary 4.4 to estimate the error in the approximation  $\theta_n(X) \approx \theta_n(X, N)$ . The error should be given by the expression

$$\text{err}_n(X, N) = \sqrt{\frac{6 \cdot X^{1/6} \cdot \theta_n(X)(1 - \theta_n(X))}{5N\kappa}} = \sqrt{\frac{6 \cdot X^{1/6}}{5N\kappa} \cdot \frac{s_n}{1 + C_n X^{-e_n}} \left(1 - \frac{s_n}{1 + C_n X^{-e_n}}\right)}.$$

In Table 6 we include the values of:  $\theta_n(X, N)$ , our model of  $\theta_n(X)$ , the error  $|\theta_n(X, N) - \theta_n(X)|$ , and the predicted standard error  $\text{err}_n(X, N)$ , for  $X = 2.6975 \cdot 10^{10}$  and  $N = 0.025 \cdot 10^9$ .

$n$	1	2	3	4	5
$\theta_n(2.6975 \cdot 10^{10}, 0.025 \cdot 10^9)$	0.44083621	0.26278969	0.05835152	0.00463836	0.00008751
$\theta_n(2.6975 \cdot 10^{10})$	0.44223400	0.26066727	0.05781814	0.00451697	0.00009141
Error	0.00139779	0.00212241	0.00053337	0.00012138	0.00000390
$\text{err}_n(2.6975 \cdot 10^{10}, 0.025 \cdot 10^9)$	0.00115688	0.00102258	0.00054367	0.00015619	0.00002227

TABLE 6. Values of:  $\theta_n(X, N)$ , our model of  $\theta_n(X)$ , the error  $|\theta_n(X, N) - \theta_n(X)|$ , and the predicted standard error  $\text{err}_n(X, N)$ , for  $X = 2.675 \cdot 10^{10}$  and  $N = 0.025 \cdot 10^9$ .

**Remark 4.7.** The BHKSSW database ([1]) also includes small databases of random samples of elliptic curves at larger heights. In particular, for each  $k \in [11, 16]$ , they calculated the Selmer rank and rank of a set  $\mathcal{L}_k$  consisting of about 100,000 curves from a uniform distribution of all curves in the height range  $[10^k, 2 \cdot 10^k)$ . We have tested  $H_A$  on these sets  $\mathcal{E}_k$  of curves of large height. In Table 7, we include the value of the moving ratio for  $\mathcal{E}_{16} \cap \mathcal{S}_n$ , the value of  $\theta_n(10^{16})$ , the error, and the predicted error  $\text{err}_n$ . The predicted error for  $n = 5$  is too large (similarly for  $n = 4$  to a lesser degree), so the sample is just too small to provide significant evidence. Otherwise, the data for  $n = 1, 2, 3$  shows that  $H_A$  and Conjecture 4.5 seem to hold even for large heights.

**Example 4.8.** As we mentioned above, Hypothesis A implies that if  $E_1, \dots, E_m$  are elliptic curves with height  $\approx X$ , then the number of curves  $E_i$  of Selmer rank  $n$  would follow a binomial distribution

$n$	1	2	3	4	5
$\#\mathcal{E}_{16} \cap \mathcal{S}_n$	42,631	27,543	7327	836	38
$\#\mathcal{E}_{16} \cap \mathcal{S}_n / \#\mathcal{E}_{16}$	0.42636116	0.27546305	0.07327879	0.00836100	0.00038004
$\theta_n(10^{16})$	0.42678631	0.27550444	0.07516196	0.00968314	0.00066148
$ \text{Error} $	0.00042515	0.00004138	0.00188317	0.00132213	0.00028144
$\text{err}_n(\mathcal{E}_{16} \cap \mathcal{S}_n)$	0.00239552	0.00269200	0.00308012	0.00338681	0.00275683

TABLE 7. Values of: the moving ratio  $\theta_n$ , our model of  $\theta_n(X)$ , the error, and the predicted standard error  $\text{err}_n(X, N)$ , for the database  $\mathcal{E}_{16}$  of height  $X \approx 10^{16}$ .

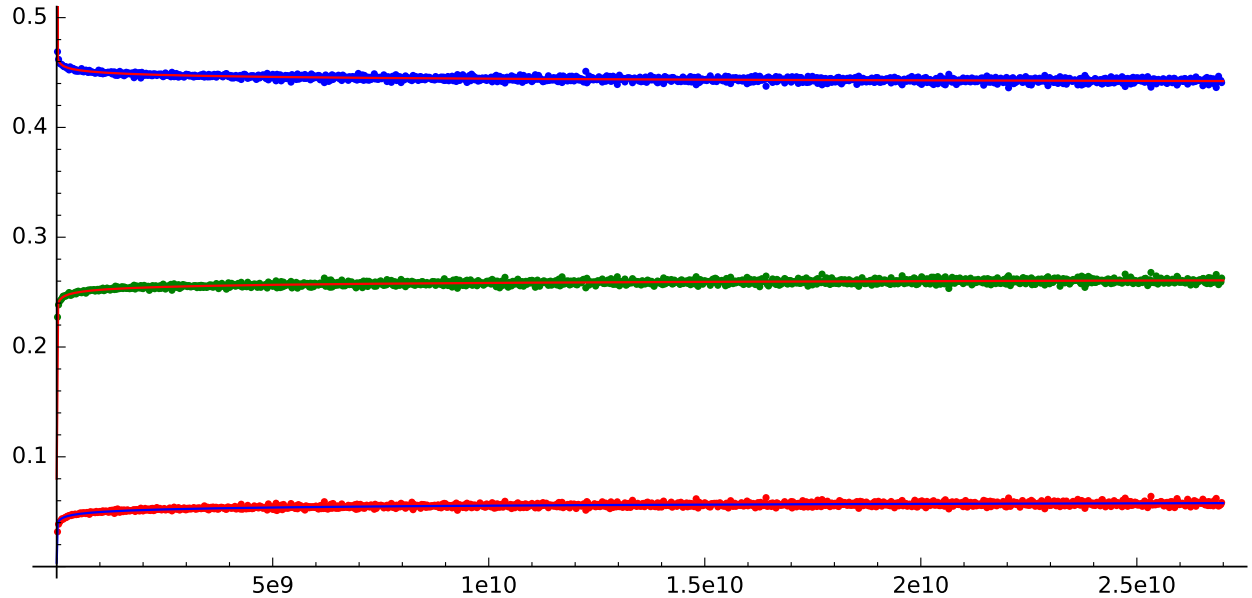


FIGURE 3. Graphs of the moving ratios  $\theta_n(X, 0.025 \cdot 10^9)$  for  $n = 1$  (blue), 2 (green), 3 (red), and the corresponding models of the form  $s_n/(1 + C_n X^{-e_n})$  (in red).

$B(m, \theta_n(X))$ . We have tested this against the BHKSSW database and the data and have always found the result to be in nice agreement with the predictions. For instance, let  $\mathcal{T}$  be the first 100,000 elliptic curves with height  $\geq 9 \cdot 10^9$ . We let  $m = 100$ , and pick 100 curves at random in  $\mathcal{T}$ , and repeat this process 10000 times. For a fixed  $n$  and for each of the 10000 trials, the distribution of the number  $0 \leq t \leq 100$  of curves with Selmer rank  $n$  would follow a binomial  $B(100, \theta_n(X))$ , where  $X$  is in the interval  $[9000573228, 9012972924]$ . We use our models  $\theta_n(X) \approx s_n/(1 + C_n X^{-e_n})$  in order to approximate values, for  $n = 1, \dots, 4$ . We obtain:

$$\theta_1(X) \approx 0.444608, \theta_2(X) \approx 0.258128, \theta_3(X) \approx 0.055273, \theta_4(X) \approx 0.003948$$

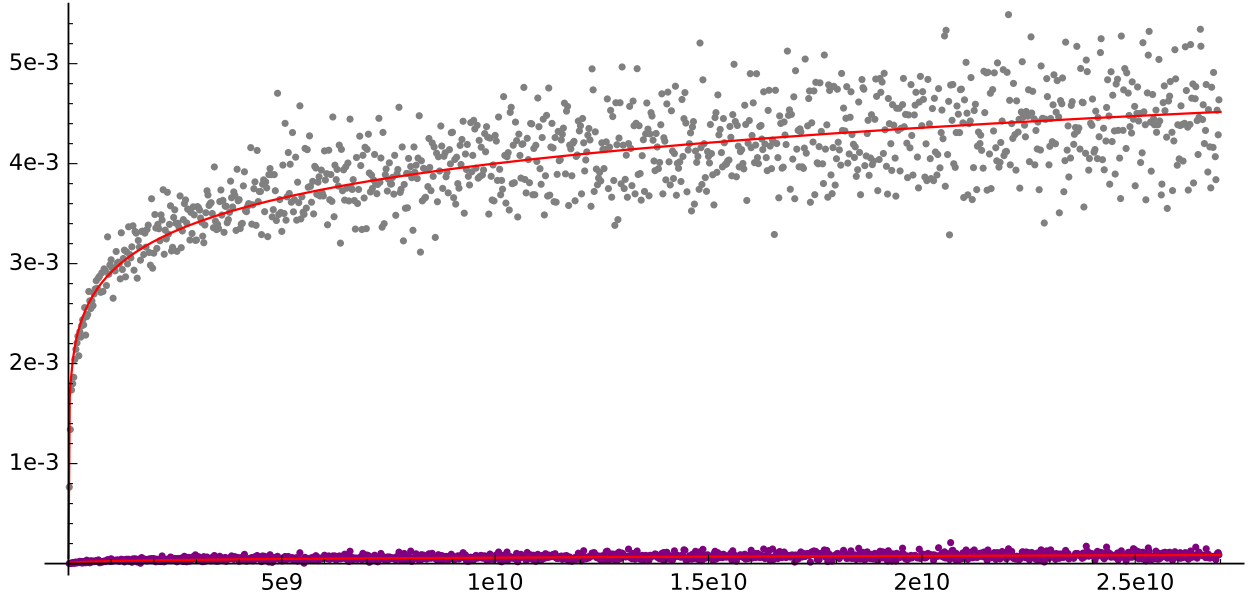


FIGURE 4. Graphs of the moving ratios  $\theta_n(X, 0.025 \cdot 10^9)$  for  $n = 4$  (gray), 5 (purple), and the corresponding models of the form  $s_n/(1 + C_n X^{-e_n})$  (in red, and blue).

for any  $X$  in the given interval. If our event of picking 100 curves follows a binomial  $B(100, \theta_n(X))$ , then it must be approximately a normal  $N(100 \cdot \theta_n(X), 100 \cdot \theta_n(X) \cdot (1 - \theta_n(X)))$ , where  $100 \cdot \theta_n(X)$  and  $100 \cdot \theta_n(X) \cdot (1 - \theta_n(X))$  are, respectively, the mean and the variance of the binomial distribution. We have plotted the result of the 10000 experiments in Figure 5, together with the normal distributions predicted by  $H_A$  and the model for  $\theta_n(X)$ .

Now, we can put our results together to estimate the number of curves of Selmer rank  $1, \dots, 5$  up to height  $X$ .

**Proposition 4.9.** *If we assume  $H_A$ , then:*

(1) *The expected value of  $\pi_{\mathcal{S}_n}(X)$  is given by*

$$\mathbb{E}(\pi_{\mathcal{S}_n}(X)) = \frac{5\kappa}{6} \int_0^X \frac{\theta_n(H)}{H^{1/6}} dH + \theta_n(X) \cdot O\left(X^{1/2}\right),$$

where  $\kappa = 2^{4/3} \cdot (\zeta(10) \cdot 3^{3/2})^{-1}$ . If in addition we assume Conjecture 4.5, then

$$\pi_{\mathcal{S}_n}(X) \approx \frac{5\kappa s_n}{6} \int_0^X \frac{1}{H^{1/6}(1 + C_n H^{-e_n})} dH + O\left(X^{1/2}\right),$$

where the constants  $s_n$ ,  $C_n$ , and  $e_n$  are given in Tables 3 and 5.

(2) *If  $X > N^2 \geq 0$ , and we assume Conjecture 4.5, then the expected value, on average, is*

$$\mathbb{E}(\pi_{\mathcal{S}_n}((X, X + N])) \approx \frac{5\kappa N \theta_n(X)}{6X^{1/6}} + O\left(\frac{N}{X^{1/2}}\right) \approx \frac{5\kappa s_n N}{6X^{1/6}(1 + C_n X^{-e_n})} + O\left(\frac{N}{X^{1/2}}\right).$$

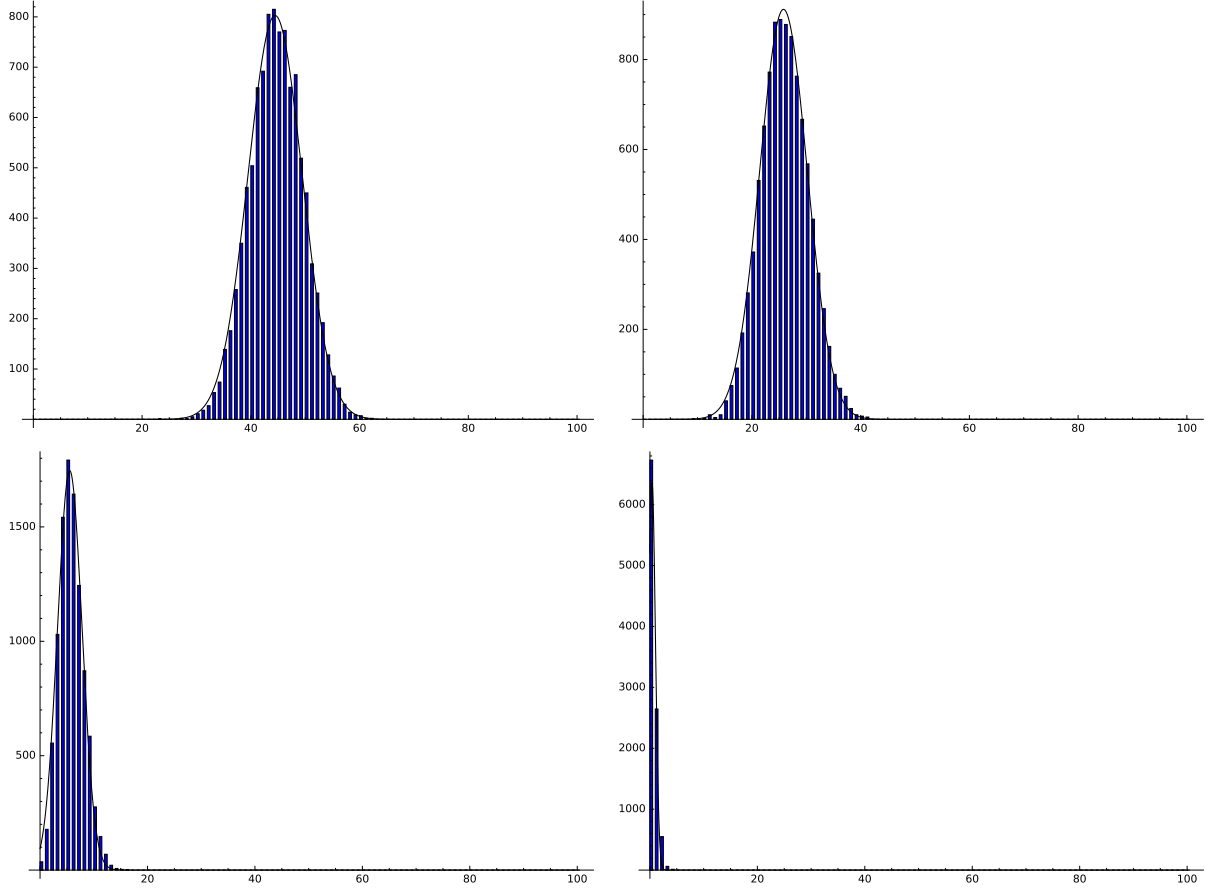


FIGURE 5. Histogram of the distribution of 10000 experiments of picking 100 elliptic curves at random of height  $\approx 9 \cdot 10^9$ , and counting the number of Selmer ranks equal to  $n = 1, 2, 3, 4$ . The graph is that of the normal distribution predicted by  $H_A$ .

*Proof.* If we assume  $H_A$ , then the expected value of  $\pi_{\mathcal{S}_n}(X) = \sum_{H=1}^X \pi_{\mathcal{S}_n}([H, H])$  is given by  $\sum_{H=1}^X \pi_{\mathcal{E}}([H, H]) \cdot \theta_n(H)$ . Thus, Corollary 3.4 and Lemma 3.6 imply that

$$\begin{aligned} \mathbb{E}(\pi_{\mathcal{S}_n}(X)) &= \sum_{H=1}^X \pi_{\mathcal{E}}([H, H]) \cdot \theta_n(H) = \frac{5\kappa}{6} \int_0^X \frac{\theta_n(H)}{H^{1/6}} dH + \theta_n(X) \cdot O(X^{1/2}) \\ &\approx \frac{5\kappa s_n}{6} \int_0^X \frac{1}{H^{1/6}(1 + C_n H^{-e_n})} dH + \theta_n(X) \cdot O(X^{1/2}), \end{aligned}$$

where the last approximation assumes Conjecture 4.5. For part (2), if  $X > N^2 \geq 0$ , then by Corollary 3.4, the expected value on average is given by

$$\begin{aligned} \mathbb{E}(\pi_{\mathcal{S}_n}((X, X+N])) &= \sum_{H=X+1}^{X+N} \pi_{\mathcal{E}}([H, H]) \cdot \theta_n(H) \approx \frac{5\kappa}{6} \sum_{H=X+1}^{X+N} \left( \frac{\theta_n(H)}{H^{1/6}} + O\left(\frac{1}{X^{1/2}}\right) \right) \\ &\approx \frac{5\kappa N \theta_n(X)}{6X^{1/6}} + O\left(\frac{N}{X^{1/2}}\right) \\ &\approx \frac{5\kappa \mathcal{S}_n N}{6X^{1/6}(1 + C_n X^{-e_n})} + O\left(\frac{N}{X^{1/2}}\right), \end{aligned}$$

as claimed.  $\square$

We have used SageMath to do numerical integration and approximation of the values of  $\pi_{\mathcal{S}_n}(X)$  using the formula of Proposition 4.9, part (a), and we have plotted the graphs against actual data from the BHKSSW database in Figures 6 and 7.

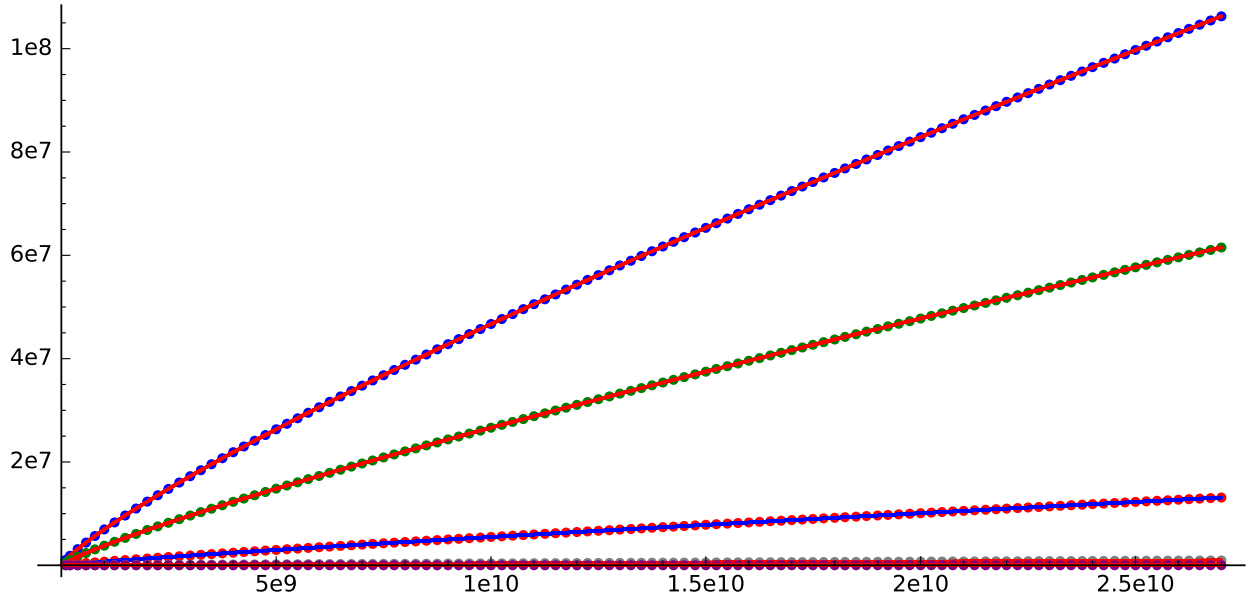


FIGURE 6. Values of  $\pi_{\mathcal{S}_n}(X)$  using the BHKSSW database are represented by dots for  $n = 1$  (blue), 2 (green), 3 (red), and the corresponding predictions from Proposition 4.9 (curves in red, except for  $n = 3$  in blue).

Finally, Proposition 4.9 will allow us to write formulas for the average 2-Selmer rank of an elliptic curve up to height  $X$ . We plot our conjectural formula in Figure 8.

**Proposition 4.10.** *Let  $\text{AvgSelRank}(X)$  be defined by*

$$\text{AvgSelRank}(X) = \frac{1}{\pi_{\mathcal{E}}(X)} \sum_{E \in \mathcal{E}(X)} \text{selrank}(E(\mathbb{Q})).$$

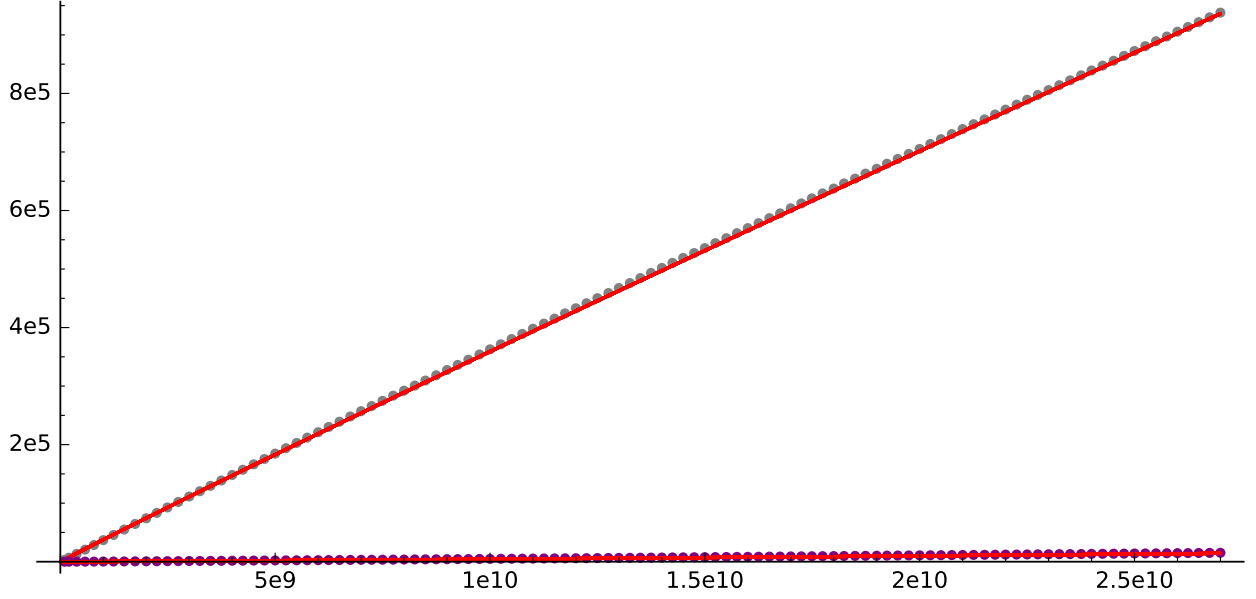


FIGURE 7. Values of  $\pi_{S_n}(X)$  using the BHKSSW database are represented by dots for  $n = 4$  (gray),  $5$  (purple), and the corresponding predictions from Proposition 4.9 (curves in red).

If we assume  $H_A$  and we assume that  $0 \leq \theta_n(X) \leq s_n$  for all  $n \geq 2$  and all  $X > 0$ , then the expected value of the average Selmer rank is given by

$$\mathbb{E}(\text{AvgSelRank}(X)) = \frac{5/6}{X^{5/6}} \int_0^X \frac{\sum_{n \geq 1} n \cdot \theta_n(H)}{H^{1/6}} dH + O(X^{-1/3}).$$

Moreover,  $\lim_{X \rightarrow \infty} \mathbb{E}(\text{AvgSelRank}(X)) = \sum_{n \geq 1} n s_n = 1.26449978 \dots$

*Proof.* In order to compute the average Selmer rank, we note that

$$\text{AvgSelRank}(X) = \frac{1}{\pi_{\mathcal{E}}(X)} \sum_{E \in \mathcal{E}(X)} \text{selrank}(E(\mathbb{Q})) = \frac{1}{\pi_{\mathcal{E}}(X)} \sum_{n \geq 1} \sum_{E \in S_n(X)} n = \frac{1}{\pi_{\mathcal{E}}(X)} \sum_{n \geq 1} n \cdot \pi_{s_n}(X).$$

Thus, by Prop. 4.9 we have that the expected value of  $\text{AvgSelRank}(X)$  is given by

$$\mathbb{E}(\text{AvgSelRank}(X)) = \frac{1}{\pi_{\mathcal{E}}(X)} \left( \frac{5\kappa}{6} \int_0^X \frac{\sum_{n \geq 1} n \cdot \theta_n(H)}{H^{1/6}} dH + \left( \sum_{n \geq 1} n \cdot \theta_n(X) \right) \cdot O(X^{1/2}) \right)$$

Let us define  $t_1 = s_1$  and  $t_n = t_1 / (2^{\frac{n(n-1)}{2}-1})$  for  $n \geq 2$ . Then, the definition of  $s_n$  implies that  $s_n \leq t_n$ , and so,

$$\sum_{n=2}^N n s_n \leq \sum_{n=2}^N n t_n \leq \sum_{n=2}^N \frac{n t_1}{2^{\frac{n(n-1)}{2}-1}}$$

for any  $N \geq 2$ . In particular,  $\sum_{n \geq 1} ns_n$  converges. Since we are assuming  $0 \leq \theta_n(X) \leq s_n$  for  $n \geq 2$ , it follows that  $\sum_{n \geq 1} n\theta_n(X)$  converges for any  $X$ , and  $\lim_{X \rightarrow \infty} \sum_{n \geq 1} n\theta_n(X) = \sum_{n \geq 1} ns_n = 1.26449978 \dots$ . Thus, if we use this, and  $\pi_{\mathcal{E}}(X) = \kappa X^{5/6} + O(X^{1/2})$  from Theorem 3.1, we obtain

$$\mathbb{E}(\text{AvgSelRank}(X)) = \frac{5/6}{X^{5/6}} \int_0^X \frac{\sum_{n \geq 1} n \cdot \theta_n(H)}{H^{1/6}} dH + O(X^{-1/3}).$$

Next, we calculate the limit of  $\mathbb{E}(\text{AvgSelRank}(X))$  as  $X \rightarrow \infty$ . Let  $\alpha = \sum_{n \geq 1} ns_n$ . Then,

$$\begin{aligned} \mathbb{E}(\text{AvgSelRank}(X)) &= \frac{5/6}{X^{5/6}} \int_0^X \frac{\sum_{n \geq 1} n \cdot \theta_n(H)}{H^{1/6}} dH + O(X^{-1/3}) \\ &= \frac{5/6}{X^{5/6}} \left( \int_0^X \frac{(\sum_{n \geq 1} n \cdot \theta_n(H) - \alpha)}{H^{1/6}} dH + \int_0^X \frac{\alpha}{H^{1/6}} dH \right) + O(X^{-1/3}). \end{aligned}$$

Now, since  $f(X) = \sum_{n \geq 1} n \cdot \theta_n(X) - \alpha$  goes to 0 as  $X \rightarrow \infty$ , it follows that  $X^{-5/6} \int_0^X f(H) H^{-1/6} dH$  also vanishes in the limit. Hence,

$$\begin{aligned} \lim_{X \rightarrow \infty} \mathbb{E}(\text{AvgSelRank}(X)) &= \frac{5/6}{X^{5/6}} \int_0^X \frac{\sum_{n \geq 1} n \cdot \theta_n(H)}{H^{1/6}} dH + O(X^{-1/3}) \\ &= \lim_{X \rightarrow \infty} \frac{5/6}{X^{5/6}} \int_0^X \frac{\alpha}{H^{1/6}} dH = \alpha = \sum_{n \geq 1} ns_n, \end{aligned}$$

as we wanted to prove. □

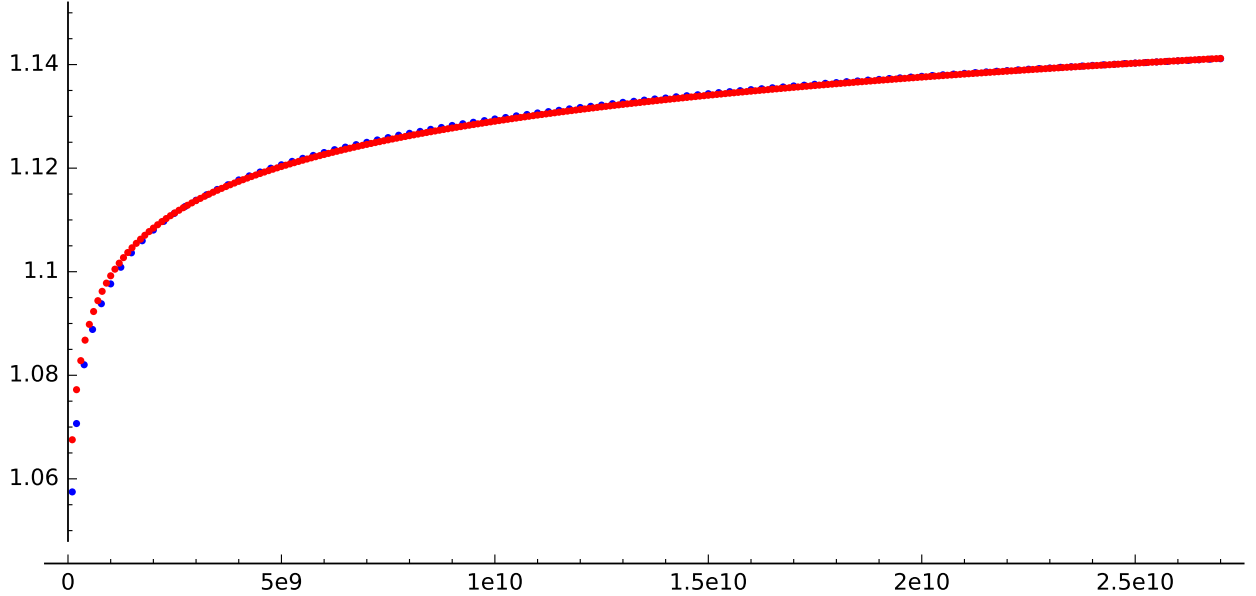


FIGURE 8. Values of  $\text{AvgSelRank}(X)$  using the BHKSSW database (blue dots), and the corresponding predictions from Proposition 4.10 (red dots).

## 5. THE PROBABILITY THAT A 2-SELMER ELEMENT IS GLOBALLY SOLVABLE

Let  $E/\mathbb{Q}$  be an elliptic curve, and let  $\text{Sel}_2(E/\mathbb{Q})$  and  $\text{III}(E/\mathbb{Q})[2]$  be, respectively, the 2-Selmer group of  $E/\mathbb{Q}$  and the 2-torsion of Sha. We would like to understand how often an element of  $\text{Sel}_2(E/\mathbb{Q})$  is in the image of  $E(\mathbb{Q})/2E(\mathbb{Q})/(E(\mathbb{Q})_{\text{tors}}/2E(\mathbb{Q})_{\text{tors}})$  under the natural injection  $\delta_E$  of the short sequence in Eq. (2) already mentioned in Section 4. Equivalently, we would like to know when an element of  $\text{Sel}_2(E/\mathbb{Q})$  reduces to a non-trivial element in the quotient  $\text{Sel}_2(E/\mathbb{Q})/(E(\mathbb{Q})/2E(\mathbb{Q})) \cong \text{III}(E/\mathbb{Q})[2]$ . Inspired by the Cohen-Lenstra heuristics for number fields, Delaunay ([8], [9]) has conjectured certain distributions of Tate-Shafarevich groups (see also Section 5 of [21] for a rich account of Delaunay's conjectures and other related works). As in the case of the results on the density of Selmer ranks discussed in Section 4, Delaunay's heuristics provide the (conjectural) limit value of the density (i.e., the total probability) of curves with a certain structure of  $\text{III}(E/\mathbb{Q})$ . However, for our purposes, we are interested in the average size of  $\text{III}(E/\mathbb{Q})[2]$  at height  $X$ , for a curve of fixed Selmer rank  $n$ . In other words, we are interested in the following conditional probability that measures the failure of the Hasse principle at a given 2-Selmer element of height  $X$ :

$$\text{pHasse}_n(X) = \text{Prob}(s \in \text{Sel}_2(E/\mathbb{Q}) \text{ is trivial in } \text{III}(E/\mathbb{Q})[2] \mid E \in \mathcal{S}_n \text{ and } \text{ht}(E/\mathbb{Q}) = X).$$

An element  $s \in \text{Sel}_2(E/\mathbb{Q})$ , in turn, can be visualized as a homogeneous space  $H \in WC(E/\mathbb{Q})$  in the Weil-Châtelet group of  $E/\mathbb{Q}$ , such that  $H$  is locally solvable everywhere, and the quantity  $\text{pHasse}_n(X)$  would be realized as the probability of  $H(\mathbb{Q})$  having a rational point (see [7]).

At this point, we could measure  $\text{pHasse}_n(E/\mathbb{Q})$ , the average failure of the Hasse principle for the 2-Selmer elements coming from a fixed elliptic curve  $E$  of height  $X$ , as usual, by

$$\frac{1}{\#\text{III}(E/\mathbb{Q})[2]} = \frac{\#(E(\mathbb{Q})/2E(\mathbb{Q}))}{\#\text{Sel}_2(E/\mathbb{Q})} = \frac{1}{2^{\text{selrank}(E(\mathbb{Q})) - \text{rank}(E(\mathbb{Q}))}}.$$

However, this ratio does not capture correctly the *probability* that a 2-Selmer element is trivial in  $\text{III}$ . Indeed, it is important to note that if  $s, s' \in \text{Sel}_2(E/\mathbb{Q})$  are two distinct elements, then the events  $s \equiv 0 \in \text{III}(E/\mathbb{Q})$  and  $s' \equiv 0 \in \text{III}(E/\mathbb{Q})$  are in general *not* independent from a probabilistic point of view. Indeed, if the 2-Selmer rank of  $E/\mathbb{Q}$  is  $n$ , then  $\text{Sel}_2(E/\mathbb{Q})$  (modulo 2-torsion contributions) has order  $2^n$ , but the size of  $\text{III}(E/\mathbb{Q})[2]$  is dictated by the classes of  $n$  generators  $s_1, \dots, s_n$  of  $\text{Sel}_2(E/\mathbb{Q})/(2\text{-torsion})$ . Thus, a better measure for  $\text{pHasse}_n(X)$  may be

$$\frac{\text{rank}_{\mathbb{Z}/2\mathbb{Z}}(E(\mathbb{Q})/2E(\mathbb{Q})) - \text{rank}_{\mathbb{Z}/2\mathbb{Z}} E(\mathbb{Q})[2]}{\text{rank}_{\mathbb{Z}/2\mathbb{Z}}(\text{Sel}_2(E/\mathbb{Q})) - \text{rank}_{\mathbb{Z}/2\mathbb{Z}} E(\mathbb{Q})[2]} = \frac{\text{rank}(E(\mathbb{Q}))}{\text{selrank}(E(\mathbb{Q}))}.$$

As it turns out, this ratio is not the correct measure either for odd Selmer rank. If we assume that  $\text{III}(E/\mathbb{Q})[2^\infty]$  is finite, then the existence of the Cassels-Tate pairing ([5])

$$\Gamma : \text{III}(E/\mathbb{Q})[2^\infty] \times \text{III}(E/\mathbb{Q})[2^\infty] \rightarrow \mathbb{Q}/\mathbb{Z},$$

which is a non-degenerate, alternating, and bilinear, implies that the  $\mathbb{F}_2$ -dimension of  $\text{III}(E/\mathbb{Q})[2]$  is always even. It follows that  $\text{rank}(E(\mathbb{Q})) \equiv \text{selrank}(E(\mathbb{Q})) \pmod{2}$ . In particular, if  $\text{selrank}(E(\mathbb{Q})) = n = 2k$  or  $1 + 2k$ , then the  $\mathbb{F}_2$ -dimension of  $\text{III}(E/\mathbb{Q})[2]$  is in fact dictated by  $2k$  classes of  $\text{Sel}_2(E/\mathbb{Q})$  (if  $n = 1$ , then  $k = 0$ , so we will assume that  $n \geq 2$  from now on in this section). Therefore, the correct way to define the failure of the Hasse principle for a given elliptic curve is as follows.



**Definition 5.1.** Let  $E/\mathbb{Q}$  be an elliptic curve of Selmer rank  $n \geq 2$ . We define the average ratio of failure of the Hasse principle of the 2-Selmer elements of  $E/\mathbb{Q}$  by

$$\text{pHasse}_n(E/\mathbb{Q}) = \begin{cases} \frac{\text{rank}(E(\mathbb{Q}))}{\text{selrank}(E(\mathbb{Q}))} & \text{if } n \text{ is even, and} \\ \frac{\text{rank}(E(\mathbb{Q})) - 1}{\text{selrank}(E(\mathbb{Q})) - 1} & \text{if } n \text{ is odd.} \end{cases}$$

We note here that, in all cases, we have

$$\text{pHasse}_n(E/\mathbb{Q}) = \frac{\text{rank}(E(\mathbb{Q})) - (n \bmod 2)}{n - (n \bmod 2)}.$$

**Remark 5.2.** The fact that the  $\mathbb{F}_2$ -dimension of  $\text{III}(E/\mathbb{Q})[2]$  is even implies that  $n = \text{selrank}(E(\mathbb{Q}))$  and  $\text{rank}(E(\mathbb{Q}))$  have the same parity. Thus, the rank of  $E(\mathbb{Q})$  is determined by  $\lfloor n/2 \rfloor$  pairs of generators  $\{(s_1, \hat{s}_1), (s_2, \hat{s}_2), \dots, (s_{\lfloor n/2 \rfloor}, \hat{s}_{\lfloor n/2 \rfloor})\}$  of  $\text{Sel}_2(E/\mathbb{Q})$  such that  $s_i \equiv 0 \in \text{III}$  if and only if  $\hat{s}_i \equiv 0 \in \text{III}$  (see Theorem 5.9 for a proof of this fact). Thus, it may be best to define

$$\text{pHasse}_n(E/\mathbb{Q}) = \frac{\frac{1}{2}(\text{rank}(E(\mathbb{Q})) - (n \bmod 2))}{\frac{1}{2}(\text{selrank}(E(\mathbb{Q})) - (n \bmod 2))}$$

but, of course, the factors of  $\frac{1}{2}$  cancel out and this definition is equivalent to the one given above. This simple remark will be crucial when computing the probability of a given Mordell-Weil rank  $r$  among curves of Selmer rank  $n$  in Theorem 5.9.

Now we are ready to state our second hypothesis of our model for the distribution of ranks.

**Hypothesis 5.3** (Hypothesis B, or  $H_B$ ). Let  $n \geq 2$  be fixed, let  $0 \leq a \leq b$ , and define

$$\text{Sel}_n([a, b]) = \bigcup_{E \in \mathcal{S}_n([a, b])} \{s_{E,1}, \dots, s_{E,n-(n \bmod 2)}\}$$

where, for a fixed elliptic curve  $E/\mathbb{Q}$ , the  $\{s_{E,1}, \dots, s_{E,n-(n \bmod 2)}\} \subseteq \text{Sel}_2(E/\mathbb{Q})$  is formed by any set of  $n - (n \bmod 2)$  generators of  $\text{Sel}_2(E/\mathbb{Q})$  as a  $\mathbb{Z}/2\mathbb{Z}$ -vector space. Let  $Y_{\text{Hasse},n,X} : \text{Sel}_n([X, X]) \rightarrow \{0, 1\}$  be the function that takes values

$$Y_{\text{Hasse},n,X}(s_E) = \begin{cases} 1 & \text{if } s_E \equiv 0 \in \text{III}(E/\mathbb{Q})[2], \\ 0 & \text{otherwise.} \end{cases}$$

There exists a function  $\rho_n(X)$ , continuous for  $X > 0$ , such that if  $\text{Sel}_n([X, X]) \neq \emptyset$ , then  $Y_{\text{Hasse},n,X}$  behaves like a random variable that follows a Bernoulli distribution with probability  $\rho_n(X)$ , and such that  $\lim_{X \rightarrow \infty} \rho_n(X) = 0$ .

Moreover:

- (1) If  $s$  and  $s'$  are Selmer elements from non-isogenous elliptic curves  $E$  and  $E'$  with heights  $X$  and  $X'$  respectively, then the random variables  $Y = Y_{\text{Hasse},n,X}(s)$  and  $Y' = Y_{\text{Hasse},n,X'}(s')$  are independent and, therefore, uncorrelated (that is,  $\text{Cov}(Y, Y') = 0$ ).
- (2) Let  $E \in \mathcal{S}_n^X$  be fixed, let  $\{(s_1, \hat{s}_1), (s_2, \hat{s}_2), \dots, (s_{\lfloor n/2 \rfloor}, \hat{s}_{\lfloor n/2 \rfloor})\}$  be a set of  $\lfloor n/2 \rfloor$  generators of  $\text{Sel}_2(E/\mathbb{Q})$ , chosen as in Remark 5.2 (see Theorem 5.9 also), and  $Y_i = Y_{\text{Hasse},n,X}(s_i)$  for  $1 \leq i \leq \lfloor n/2 \rfloor$ . We also fix an integer  $k$  with  $1 \leq k \leq \lfloor n/2 \rfloor$ . Then, the expected value of the

product of  $k$  distinct random variables in the set  $\{Y_i\}$  depends on  $k$  but it is independent of the chosen curve  $E \in \mathcal{S}_n^X$  and is independent of the choice of indices. In other words,

$$\mathbb{E}_k^n(X) = \mathbb{E}(Y_{i_1} Y_{i_2} \cdots Y_{i_k}),$$

is independent of the choice of  $E \in \mathcal{S}_n^X$  and the indices  $1 \leq i_1 < i_2 < \cdots < i_k \leq \lfloor n/2 \rfloor$ .

**Remark 5.4.** As mentioned in Remark 4.2, when considering large sample sets, we can replace “non-isogenous” (in part (1) of  $H_B$ ) by “non-isomorphic”, and we will also neglect here the (small) error introduced by possibly considering isogenous curves in our sample sets.

**Definition 5.5.** When a set of random variables  $Y_1, \dots, Y_m$  satisfy a condition as in part (2) of  $H_B$ , we shall say that they are equicorrelated.

**Remark 5.6.** The equicorrelation condition of  $H_B$ , part (2), does not add any conditions at all when  $n = 1, 2, 3$ . When  $n = 4, 5$ , equicorrelation simply says that  $\mathbb{E}(Y_1) = \mathbb{E}(Y_2)$  which is already implied by the assumption that  $Y_1$  and  $Y_2$  follow the same Bernoulli distribution (so in fact  $\mathbb{E}(Y_1) = \mathbb{E}(Y_2) = \rho_n(X)$ ). However, the equicorrelation does add new information about the random variables  $\{Y_i\}$  for  $n \geq 6$ . For instance, when  $n = 6$ , it says that

$$\mathbb{E}(Y_1 Y_2) = \mathbb{E}(Y_1 Y_3) = \mathbb{E}(Y_2 Y_3).$$

When  $n = 8$ , it says that

$$\mathbb{E}(Y_1 Y_2) = \mathbb{E}(Y_1 Y_3) = \mathbb{E}(Y_1 Y_4) = \mathbb{E}(Y_2 Y_3) = \mathbb{E}(Y_2 Y_4) = \mathbb{E}(Y_3 Y_4),$$

and also

$$\mathbb{E}(Y_1 Y_2 Y_3) = \mathbb{E}(Y_1 Y_2 Y_4) = \mathbb{E}(Y_1 Y_3 Y_4) = \mathbb{E}(Y_2 Y_3 Y_4).$$

The following two results describe the effects of equicorrelation on the covariance of the random variables.

**Lemma 5.7.** Let  $Z, Z', W, W'$  be random variables such that  $\mathbb{E}(Z) = \mathbb{E}(Z')$ ,  $\mathbb{E}(W) = \mathbb{E}(W')$ . Then,  $\text{Cov}(Z, W) = \text{Cov}(Z', W')$  if and only if  $\mathbb{E}(ZW) = \mathbb{E}(Z'W')$ , if and only if  $\mathbb{E}((1 - Z)W) = \mathbb{E}((1 - Z')W')$ .

*Proof.* By definition  $\mathbb{E}(ZW) = \mathbb{E}(Z)\mathbb{E}(W) + \text{Cov}(Z, W)$ . Thus,

$$\begin{aligned} \mathbb{E}(ZW) - \mathbb{E}(Z'W') &= \mathbb{E}(Z)\mathbb{E}(W) + \text{Cov}(Z, W) - (\mathbb{E}(Z')\mathbb{E}(W') + \text{Cov}(Z', W')) \\ &= \mathbb{E}(Z)\mathbb{E}(W) - \mathbb{E}(Z')\mathbb{E}(W') + \text{Cov}(Z, W) - \text{Cov}(Z', W') \\ &= \text{Cov}(Z, W) - \text{Cov}(Z', W'). \end{aligned}$$

Thus,  $\text{Cov}(Z, W) = \text{Cov}(Z', W')$  if and only if  $\mathbb{E}(ZW) = \mathbb{E}(Z'W')$ . Similarly,

$$\begin{aligned} &\mathbb{E}((1 - Z)W) - \mathbb{E}((1 - Z')W') \\ &= \mathbb{E}(1 - Z)\mathbb{E}(W) + \text{Cov}(1 - Z, W) - (\mathbb{E}(1 - Z')\mathbb{E}(W') + \text{Cov}(1 - Z', W')) \\ &= (1 - \mathbb{E}(Z))\mathbb{E}(W) - (1 - \mathbb{E}(Z'))\mathbb{E}(W') - \text{Cov}(Z, W) + \text{Cov}(Z', W') \\ &= -\text{Cov}(Z, W) + \text{Cov}(Z', W'), \end{aligned}$$

as claimed, where we have used the fact that  $\text{Cov}(a + bX, Y) = b \text{Cov}(X, Y)$ , for any constants  $a, b$  and random variables  $X, Y$ .  $\square$

**Proposition 5.8.** *Assume  $H_B$ , and let  $Y_i = Y_{\text{Hasse},n,X}(s_{E,i})$  for  $1 \leq i \leq \lfloor n/2 \rfloor$  be the random variables defined in  $H_B$ , part (2), that are assumed to be equicorrelated. Let  $1 \leq s, t \leq m$  with  $s + t \leq m$ . Then, there is a function  $C_{s,t}^n(X)$  such that*

$$C_{s,t}^n(X) = \text{Cov}(Y_{i_1} Y_{i_2} \cdots Y_{i_s}, Y_{k_1} Y_{k_2} \cdots Y_{k_t})$$

for any sets of indices  $1 \leq i_1 < i_2 < \cdots < i_s \leq m$  and  $1 \leq k_1 < k_2 < \cdots < k_t \leq m$  with  $\{i_u\} \cap \{k_v\} = \emptyset$ .

*Proof.* Let  $1 \leq i_1 < i_2 < \cdots < i_s \leq m$  and  $1 \leq k_1 < k_2 < \cdots < k_t \leq m$  with  $\{i_u\} \cap \{k_v\} = \emptyset$ , and let  $1 \leq i'_1 < i'_2 < \cdots < i'_s \leq m$  and  $1 \leq k'_1 < k'_2 < \cdots < k'_t \leq m$  with  $\{i'_u\} \cap \{k'_v\} = \emptyset$  be another set of such indices. Since the random variables are equicorrelated, we have  $\mathbb{E}_s^n(X) = \mathbb{E}(Y_{i_1} Y_{i_2} \cdots Y_{i_s}) = \mathbb{E}(Y_{i'_1} Y_{i'_2} \cdots Y_{i'_s})$  and similarly  $\mathbb{E}_t^n(X) = \mathbb{E}(Y_{k_1} Y_{k_2} \cdots Y_{k_t}) = \mathbb{E}(Y_{k'_1} Y_{k'_2} \cdots Y_{k'_t})$  and also

$$\mathbb{E}_{s+t}^n(X) = \mathbb{E}(Y_{i_1} Y_{i_2} \cdots Y_{i_s} Y_{k_1} Y_{k_2} \cdots Y_{k_t}) = \mathbb{E}(Y_{i'_1} Y_{i'_2} \cdots Y_{i'_s} Y_{k'_1} Y_{k'_2} \cdots Y_{k'_t}).$$

Then, we can apply Lemma 5.7 with  $Z = Y_{i_1} Y_{i_2} \cdots Y_{i_s}$ ,  $W = Y_{k_1} Y_{k_2} \cdots Y_{k_t}$ ,  $Z' = Y_{i'_1} Y_{i'_2} \cdots Y_{i'_s}$ , and  $W' = Y_{k'_1} Y_{k'_2} \cdots Y_{k'_t}$ , to obtain the equality of the covariance terms (and the covariance does not depend on the choice of curve  $E \in \mathcal{S}_n^X$  by  $H_B$ ). Thus, the covariance is independent of the chosen sets of  $s$  and  $t$  distinct random variables in  $\{Y_i\}$ , and in fact it only depends on  $n$ ,  $s$ ,  $t$ , and  $X$ .  $\square$

Hypothesis B asserts that  $Y_{\text{Hasse},n,X} \sim B(1, \rho_n(X))$ , i.e.,  $Y_{\text{Hasse},n,X}$  follows a binomial distribution with one trial. Now we want to reconstruct the distribution of the rank of  $E \in \mathcal{S}_n^X$  from that of  $Y_{\text{Hasse},n,X}$ .

**Theorem 5.9.** *Let  $n \geq 1$  be fixed, assume  $H_B$ , and let  $R_n = \{0, 1, \dots, \lfloor n/2 \rfloor\}$ . Then, the function  $\text{rank}_{n,X} : \mathcal{S}_n^X \rightarrow R_n$  given by  $\text{rank}_{n,X}(E) = (\text{rank}(E(\mathbb{Q})) - (n \bmod 2))/2$  is the random variable  $Y_1 + \cdots + Y_{\lfloor n/2 \rfloor}$ , where  $\{Y_i\}_{i=1}^{\lfloor n/2 \rfloor}$  is a set of  $\lfloor n/2 \rfloor$  samples of  $Y_{\text{Hasse},n,X}$ . In particular:*

- (1)  $\text{rank}_{n,X}(E) = 0$  for  $n = 1$ , for any  $E \in \mathcal{S}_1(X)$ .
- (2) If  $n \geq 2$ , The expected value of  $\text{rank}_{n,X}$  is  $\lfloor n/2 \rfloor \cdot \rho_n(X)$  and variance

$$\text{Var}(\text{rank}_{n,X}) = \lfloor n/2 \rfloor \cdot (\rho_n(X)(1 - \rho_n(X)) + (\lfloor n/2 \rfloor - 1) \cdot C_{1,1}^n(X)),$$

where  $C_{1,1}^n(X) = \text{Cov}(Y_i, Y_j)$  is the covariance function of any two random variables given by Proposition 5.8.

- (3) If the random variables  $\{Y_i\}$  are independent (resp. approximately uncorrelated, i.e., if  $C_{1,1}^n(X) \approx 0$ ), then  $\text{rank}_{n,X}$  follows (resp. approximately) a binomial distribution of the form  $B(\lfloor n/2 \rfloor, \rho_n(X))$ , with expected value  $\lfloor n/2 \rfloor \cdot \rho_n(X)$  and variance  $\lfloor n/2 \rfloor \cdot \rho_n(X)(1 - \rho_n(X))$ .
- (4) If  $E$  and  $E'$  are non-isogenous curves of Selmer rank  $n$  and heights  $X$  and  $X'$ , respectively, then the random variables  $\text{rank}_{n,X}(E)$  and  $\text{rank}_{n,X'}(E')$  are independent.

*Proof.* For part (1), note that we are assuming the finiteness of  $\text{III}(E/\mathbb{Q})[2^\infty]$  and therefore if we have  $\text{selrank}(E) = 1$ , then  $\text{rank}(E) = 1$  as well, since  $\text{selrank}(E) \equiv \text{rank}(E) \pmod{2}$ . Thus,  $\text{rank}_{n,X}(E) = 0$ . For the rest of the proof, let us now assume that  $n \geq 2$ , and first assume that  $n$  is even,  $n = 2k$ . Let  $E/\mathbb{Q}$  be an elliptic curve of Selmer rank  $n$ , and let  $s_1$  be an arbitrary element of  $\text{Sel}_2(E/\mathbb{Q})/(2\text{-torsion})$ . According to  $H_B$ , the element  $s_1$  is trivial in  $\text{III}$  with probability  $\rho_n(X)$ . We distinguish two cases:

- If  $s_1 \equiv 0 \in \text{III}$ , then  $\dim_{\mathbb{F}_2}(\text{III}(E/\mathbb{Q})[2])$  is now at most  $2k - 2$ , and this means that there exists a Selmer element  $\hat{s}_1$ , linearly independent from  $s_1$ , such that  $\hat{s}_1 \equiv 0 \in \text{III}$  as well.

- Otherwise, if  $s_1$  represents a non-trivial element in  $\text{III}$ , and if  $\Gamma : \text{III}(E/\mathbb{Q})[2] \times \text{III}(E/\mathbb{Q})[2] \rightarrow \mathbb{F}_2$  is the Cassels-Tate (non-degenerate, alternating, and bilinear) pairing, then we can choose a non-trivial element  $[\hat{s}_1] \in \text{III}(E/\mathbb{Q})[2]$  such that  $\Gamma([s_1], [\hat{s}_1]) = 1$ . In particular,  $[\hat{s}_1]$  is linearly independent of the class of  $s_1$  in  $\text{III}$ , and therefore if  $\hat{s}_1$  is now any Selmer element representing the same class  $[\hat{s}_1]$  of  $\text{III}$ , then  $\hat{s}_1$  and  $s_1$  are also linearly independent in  $\text{Sel}_2$ .

In either case, we have found a pair  $(s_1, \hat{s}_1)$  of Selmer elements with  $Y_{\text{Hasse}, n, X}(s_1) = Y_{\text{Hasse}, n, X}(\hat{s}_1)$ , and the pair is trivial in  $\text{III}$  with probability  $\rho_n(X)$ . We can continue this process to find pairs  $(s_1, \hat{s}_1), \dots, (s_k, \hat{s}_k)$  such that both elements in each pair are trivial in  $\text{III}$  with probability  $\rho_n(X)$ , by  $H_B$ . Therefore,

$$\frac{1}{2} \text{rank}(E(\mathbb{Q})) = \sum_{i=1}^k Y_{\text{Hasse}, n, X}(s_i) = \sum_{i=1}^k Y_{\text{Hasse}, n, X}(\hat{s}_i)$$

where  $k = n/2 = \lfloor n/2 \rfloor$ .

Now let  $n = 1 + 2k$  be odd. The proof is analogous, except that if  $\dim_{\mathbb{F}_2} \text{Sel}_2(E/\mathbb{Q})$  is odd, and  $\dim_{\mathbb{F}_2}(\text{III}(E/\mathbb{Q})[2])$  must be even, then there is automatically a Selmer element  $s_0$  that is trivial in  $\text{III}$ . Now we can proceed as above to find pairs  $(s_1, \hat{s}_1), \dots, (s_k, \hat{s}_k)$  such that each pair is trivial in  $\text{III}$  with probability  $\rho_n(X)$ , by  $H_B$ . Therefore,

$$\frac{\text{rank}(E(\mathbb{Q})) - 1}{2} = \sum_{i=1}^k Y_{\text{Hasse}, n, X}(s_i) = \sum_{i=1}^k Y_{\text{Hasse}, n, X}(\hat{s}_i)$$

where  $k = (n - 1)/2 = \lfloor n/2 \rfloor$ .

For (2), we no longer assume that the random variables  $Y_i$  are independent. Nonetheless, we can calculate the expected value of  $\text{rank}_{n, X}$ .

$$\mathbb{E}(\text{rank}_{n, X}(E(\mathbb{Q}))) = \mathbb{E} \left( \sum_{i=1}^{\lfloor n/2 \rfloor} Y_{\text{Hasse}, n, X}(s_i) \right) = \sum_{i=1}^{\lfloor n/2 \rfloor} \mathbb{E}(Y_{\text{Hasse}, n, X}(s_i)) = \lfloor n/2 \rfloor \cdot \rho_n(X),$$

since each  $Y_{\text{Hasse}, n, X}(s_i) \sim B(1, \rho_n(X))$  by Hypothesis B. Let us now calculate the variance of  $\text{rank}_{n, X} = \sum Y_i$ , where  $Y_i = Y_{\text{Hasse}, n, X}(s_i)$ .

$$\begin{aligned} \text{Var}(\text{rank}_{n, X}(E(\mathbb{Q}))) &= \text{Var} \left( \sum_{i=1}^{\lfloor n/2 \rfloor} Y_i \right) \\ &= \sum_{i=1}^{\lfloor n/2 \rfloor} \text{Var}(Y_i) + 2 \cdot \left( \sum_{1 \leq i < j \leq \lfloor n/2 \rfloor} \text{Cov}(Y_i, Y_j) \right) \\ &= \sum_{i=1}^{\lfloor n/2 \rfloor} \text{Var}(Y_i) + 2 \cdot \left( \frac{\lfloor n/2 \rfloor (\lfloor n/2 \rfloor - 1)}{2} \right) \cdot C_{1,1}^n(X) \\ &= \lfloor n/2 \rfloor \cdot \rho_n(X)(1 - \rho_n(X)) + \lfloor n/2 \rfloor (\lfloor n/2 \rfloor - 1) \cdot C_{1,1}^n(X) \\ &= \lfloor n/2 \rfloor \cdot (\rho_n(X)(1 - \rho_n(X)) + (\lfloor n/2 \rfloor - 1) \cdot C_{1,1}^n(X)), \end{aligned}$$

where we have used the properties of the variance, and the fact that for any  $i \neq j$ , we have  $\text{Cov}(Y_i, Y_j) = C_{1,1}^n(X)$  for all  $i \neq j$  by Proposition 5.8.

In particular, if the random variables  $Y_i = Y_{\text{Hasse},n,X}(s_i)$  were independent samples of a Bernoulli distribution (or similarly if  $C_{1,1}^n(X) \approx 0$ ), then  $\text{rank}_{n,X} = \sum Y_i$  would follow a binomial distribution  $B(\lfloor n/2 \rfloor, \rho_n(X))$ . This proves (3).

For (4), we write  $\text{rank}_{n,X}(E) = \sum Y_i$  and  $\text{rank}_{n,X'}(E') = \sum Y'_i$ . Since the random variables  $\{Y_i\}$  and  $\{Y'_i\}$  arise from non-isogenous elliptic curves, part (1) of  $H_B$  implies that  $Y_i$  and  $Y'_j$  are independent (that is,  $\mathbb{E}(Y_i Y'_j) = \mathbb{E}(Y_i) \mathbb{E}(Y'_j)$ ), for any  $1 \leq i, j \leq \lfloor n/2 \rfloor$ . Thus,  $Y = \sum Y_i$  and  $Y' = \sum Y'_i$  are also independent:

$$\begin{aligned} \mathbb{E}(YY') &= \mathbb{E}\left(\left(\sum_i Y_i\right)\left(\sum_j Y'_j\right)\right) = \mathbb{E}\left(\sum_{i,j} Y_i Y'_j\right) = \sum_{i,j} \mathbb{E}(Y_i Y'_j) = \sum_{i,j} \mathbb{E}(Y_i) \mathbb{E}(Y'_j) \\ &= \left(\sum_i \mathbb{E}(Y_i)\right)\left(\sum_j \mathbb{E}(Y'_j)\right) = \left(\mathbb{E}\left(\sum_i Y_i\right)\right)\left(\mathbb{E}\left(\sum_j Y'_j\right)\right) = \mathbb{E}(Y) \mathbb{E}(Y'). \end{aligned}$$

This completes the proof of (4) and of the theorem.  $\square$

Using Theorem 5.9, we shall describe the average rank and distribution of curves by Mordell-Weil rank in a sample set of curves of Selmer rank  $n$ .

**Corollary 5.10.** *Let  $E_1, \dots, E_m$  be non-isomorphic elliptic curves chosen at random of Selmer rank  $n$  and heights  $X_1, \dots, X_m$ . Then, the expected value of the average rank is*

$$\mathbb{E}\left(\frac{1}{m} \sum_{i=1}^m \text{rank}(E_i(\mathbb{Q}))\right) = (n \bmod 2) + \frac{2\lfloor n/2 \rfloor}{m} \sum_{i=1}^m \rho_n(X_i)$$

with standard error given by  $\frac{1}{m} \sqrt{\lfloor n/2 \rfloor \sum_{i=1}^m (\rho_n(X_i)(1 - \rho_n(X_i)) + (\lfloor n/2 \rfloor - 1) C_{1,1}^n(X_i))}$ .

*Proof.* Let  $E_1, \dots, E_m$  be as in the statement (we shall assume that the curves are non-isogenous; see Remark 5.4). Then, Theorem 5.9 gives us the expected value and variance of  $\text{rank}_{n,X_i}(E(\mathbb{Q})) = (\text{rank}(E_i(\mathbb{Q})) - (n \bmod 2))/2$  and therefore we can compute the expected value.

$$\mathbb{E}\left(\frac{1}{m} \sum_{i=1}^m \frac{\text{rank}(E_i(\mathbb{Q})) - (n \bmod 2)}{2}\right) = \frac{1}{m} \sum_{i=1}^m \mathbb{E}(\text{rank}_{n,X_i}(E(\mathbb{Q}))) = \frac{1}{m} \sum_{i=1}^m \lfloor n/2 \rfloor \rho_n(X_i).$$

Thus,  $\mathbb{E}(\frac{1}{m} \sum_{i=1}^m \text{rank}(E(\mathbb{Q}))) = (n \bmod 2) + \frac{2\lfloor n/2 \rfloor}{m} \sum_{i=1}^m \rho_n(X_i)$ , as claimed. Next, Theorem 5.9, part (4), shows that the random variables  $Z_i = \text{rank}_{n,X_i}(E(\mathbb{Q}))$  are independent because the curves  $\{E_i\}$  are non-isomorphic. In particular, the covariance  $\text{Cov}(Z_i, Z_j) = 0$  for all  $i \neq j$ , and it follows that  $\text{Var}(Z_i + Z_j) = \text{Var}(Z_i) + \text{Var}(Z_j) + 2 \text{Cov}(Z_i, Z_j) = \text{Var}(Z_i) + \text{Var}(Z_j)$ . Hence, we can compute the variance as follows:

$$\begin{aligned} \text{Var}\left(\frac{1}{m} \sum_{i=1}^m \text{rank}_{n,X_i}(E_i)\right) &= \frac{1}{m^2} \sum_{i=1}^m \text{Var}(\text{rank}_{n,X_i}(E_i)) \\ &= \frac{1}{m^2} \sum_{i=1}^m \lfloor n/2 \rfloor \cdot (\rho_n(X_i)(1 - \rho_n(X_i)) + (\lfloor n/2 \rfloor - 1) \cdot C_{1,1}^n(X_i)), \end{aligned}$$

and therefore the standard error is given by

$$\sqrt{\frac{\lfloor n/2 \rfloor}{m^2} \sum_{i=1}^m (\rho_n(X_i)(1 - \rho_n(X_i)) + (\lfloor n/2 \rfloor - 1)C_{1,1}^n(X_i))}$$

as desired.  $\square$

Before we go on to describe the probability of a given Mordell-Weil rank, we need a result on equicorrelated random variables.

**Lemma 5.11.** *Suppose that the random variables  $\{Y_i\}_{i=1}^n$  are equicorrelated. Then:*

(1) *For any  $1 \leq m \leq n$ , and any indices  $1 \leq i_1 < \dots < i_m \leq n$  and  $1 \leq i'_1 < \dots < i'_m \leq n$ ,*

$$\mathbb{E}((1 - Y_{i_1}) \dots (1 - Y_{i_m})) = \mathbb{E}((1 - Y_{i'_1}) \dots (1 - Y_{i'_m})).$$

(2) *If  $X$  and  $\{Y_i\}$  are all distinct equicorrelated random variables, and  $1 \leq m \leq n$ , then:*

$$\text{Cov}(X, (1 - Y_1)(1 - Y_2) \dots (1 - Y_m)) = \sum_{i=1}^m (-1)^i \binom{m}{i} \text{Cov}\left(X, \prod_{k=1}^i Y_k\right).$$

*Proof.* Part (1) can be easily shown via induction on  $m$ , where the induction step was essentially proved in Lemma 5.7. For part (2), we note that  $\text{Cov}(X, (1 - Y_1)(1 - Y_2) \dots (1 - Y_m))$  equals

$$\begin{aligned} &= \text{Cov}\left(X, 1 - \left(\sum_i Y_i\right) + \left(\sum_{i \neq j} Y_i Y_j\right) + \dots + (-1)^m \prod_i Y_i\right) \\ &= -\sum_i \text{Cov}(X, Y_i) + \sum_{i \neq j} \text{Cov}(X, Y_i Y_j) + \dots + (-1)^m \text{Cov}\left(X, \prod_i Y_i\right) \\ &= \sum_{i=1}^m (-1)^i \binom{m}{i} \text{Cov}\left(X, \prod_{k=1}^i Y_k\right), \end{aligned}$$

where we have used  $\text{Cov}(X, \prod_{s=1}^t Y_{i_s}) = \text{Cov}(X, Y_1 \dots Y_t)$  for any indices  $1 \leq i_1 < \dots < i_t \leq m$  by Proposition 5.8.  $\square$

**Remark 5.12.** Let us introduce some more notation to simplify our formulas. By Lemmas 5.7 and 5.11, if  $Y_1, \dots, Y_m$  are distinct equicorrelated random variables, and  $1 \leq s, t$  with  $s + t \leq m$ , then the value of

$$(3) \quad \mathbb{E}_{s,t} = \mathbb{E}(Y_{i_1} \dots Y_{i_s} (1 - Y_{i_{s+1}}) \dots (1 - Y_{i_{s+t}})),$$

is independent for any set of  $s + t$  distinct indices  $\{i_k\}_{k=1}^{s+t} \subseteq \{1, \dots, m\}$ . When the random variables  $Y_1, \dots, Y_{\lfloor n/2 \rfloor}$  are the ones given by Theorem 5.9, we will write  $\mathbb{E}_{s,t}^n(X) = \mathbb{E}_{s,t}$ , or simply  $\mathbb{E}_{s,t}^n$ , to indicate the expected value of a product of random variables as in Equation (3) above (which extends the notation  $\mathbb{E}_k^n(X) = \mathbb{E}(Y_1 \dots Y_k)$  of  $H_B$ ). We also write  $\mathbb{E}_{0,0}^1(X) = 1$ . The following lemma gives recursive formulas to compute any expected value  $\mathbb{E}_{s,t}^n$ .

**Lemma 5.13.** *Let  $Y_1, \dots, Y_{\lfloor n/2 \rfloor}$  be the equicorrelated random variables given by Theorem 5.9, let  $0 \leq s, t$  with  $s + t \leq \lfloor n/2 \rfloor$ , and let  $C_{s,t}^n(X)$  be the covariance coefficient of Proposition 5.8. Then, with notation as in Remark 5.12, we have identities*

- (1)  $\mathbb{E}_{1,0}^n = \rho_n(X)$  and  $\mathbb{E}_{0,1}^n = 1 - \mathbb{E}_{1,0}^n = 1 - \rho_n(X)$ .
- (2) If  $s \geq 1$ , then  $\mathbb{E}_{s,0}^n = \mathbb{E}_{s-1,0}^n \cdot \mathbb{E}_{1,0}^n + C_{s-1,1}^n(X)$ .
- (3) If  $t \geq 1$ , then  $\mathbb{E}_{0,t}^n = \mathbb{E}_{0,t-1}^n \cdot \mathbb{E}_{0,1}^n - \sum_{i=1}^{t-1} (-1)^i \binom{t-1}{i} C_{1,i}^n(X)$ .
- (4)  $\mathbb{E}_{s,t}^n = \mathbb{E}_{s,0}^n \cdot \mathbb{E}_{0,t}^n + \sum_{i=1}^t (-1)^i \binom{t}{i} C_{s,i}^n(X)$ .

*Proof.* (1)  $\mathbb{E}_{1,0}^n = \mathbb{E}(Y_1) = \rho_n(X)$  and  $\mathbb{E}_{0,1}^n = \mathbb{E}(1 - Y_1) = 1 - \mathbb{E}_{1,0}^n$ .

- (2) If  $s \geq 1$ , then  $\mathbb{E}_{s,0}^n = \mathbb{E}(Y_1 \cdots Y_{s-1}) \mathbb{E}(Y_s) + \text{Cov}(Y_1 \cdots Y_{s-1}, Y_s) = \mathbb{E}_{s-1,0}^n \cdot \mathbb{E}_{1,0}^n + C_{s-1,1}^n(X)$ .
- (3) If  $t \geq 1$ , then

$$\mathbb{E}_{0,t}^n = \mathbb{E}((1 - Y_1) \cdots (1 - Y_{t-1})) \mathbb{E}(1 - Y_t) + \text{Cov}((1 - Y_1) \cdots (1 - Y_{t-1}), 1 - Y_t)$$

and the covariance term equals  $-\text{Cov}((1 - Y_1) \cdots (1 - Y_{t-1}), Y_t)$  which in turn is

$$-\sum_{i=1}^{t-1} (-1)^i \binom{t-1}{i} \text{Cov}\left(Y_t, \prod_{k=1}^i Y_k\right) = -\sum_{i=1}^{t-1} (-1)^i \binom{t-1}{i} C_{1,i}^n(X)$$

by Lemma 5.11.

- (4)  $\mathbb{E}_{s,t}^n = \mathbb{E}_{s,0}^n \cdot \mathbb{E}_{0,t}^n + \text{Cov}(Y_1 \cdots Y_s, (1 - Y_1) \cdots (1 - Y_t))$  and, by Lemma 5.11, the covariance term equals  $\sum_{i=1}^t (-1)^i \binom{t}{i} C_{s,i}^n(X)$  as claimed.  $\square$

**Corollary 5.14.** *Let us assume  $H_B$ . Let  $n \geq 2$  be fixed, let  $0 \leq j \leq \lfloor n/2 \rfloor$ , and let  $E/\mathbb{Q}$  be an elliptic curve of Selmer rank  $n$  and height  $X$ . Then:*

- (1) *The random variable  $Y_{\text{rk}=n-2j}(E(\mathbb{Q})) : \mathcal{S}_n^X \rightarrow \{0, 1\}$  given by*

$$Y_{\text{rk}=n-2j}(E(\mathbb{Q})) = \begin{cases} 1 & \text{if } \text{rank}(E(\mathbb{Q})) = n - 2j, \\ 0 & \text{otherwise,} \end{cases}$$

*is given by*

$$Y_{\text{rk}=n-2j} = \sum_{1 \leq k_1 < \cdots < k_{m(j)} \leq \lfloor n/2 \rfloor} Y_{k_1} \cdot Y_{k_2} \cdots Y_{k_{m(j)}} \cdot \frac{\prod_{i=1}^{\lfloor n/2 \rfloor} (1 - Y_i)}{(1 - Y_{k_1})(1 - Y_{k_2}) \cdots (1 - Y_{k_{m(j)}})}$$

*where  $m(j) = \lfloor n/2 \rfloor - j$ , and the random variables  $\{Y_i\}$  are as given by Theorem 5.9, such that  $\text{rank}_{n,X} = \sum Y_i$ .*

- (2) *The expected value of  $Y_{\text{rk}=n-2j}$ , using the notation of Remark 5.12 and Lemma 5.13, is given by*

$$\binom{\lfloor n/2 \rfloor}{j} \cdot \mathbb{E}_{m(j),j}^n(X),$$

*where  $\mathbb{E}_{m(j),j}^n(X)$  can be calculated recursively using the formulae of Lemma 5.13.*

- (3) *Further, if we assume that the random variables  $Y_i$  are independent (resp. assume that they are approximately uncorrelated, i.e.,  $C_{s,t}^n(X) \approx 0$  for any  $s, t \geq 0$ ), then the expected value of  $Y_{\text{rk}=n-2j}$  is given by (resp. approximately given by)*

$$\text{Prob}(\text{rank}(E(\mathbb{Q})) = n - 2j \mid E \in \mathcal{S}_n^X) \cong \binom{\lfloor n/2 \rfloor}{j} \rho_n(X)^{\lfloor n/2 \rfloor - j} (1 - \rho_n(X))^j.$$

*Proof.* Let  $\{Y_i\}$  be the random variables given by Theorem 5.9, such that  $\text{rank}_{n,X} = \sum Y_i$ . It follows that  $\text{rank}_{n,X} = n - 2j$  if and only if there are exactly  $m(j)$  elements of  $\{s_1, \dots, s_{\lfloor n/2 \rfloor}\}$  that are trivial in  $\text{III}(E/\mathbb{Q})[2]$ , if and only if there are exactly  $m(j)$  indices  $1 \leq k_1 < \dots < k_{m(j)} \leq \lfloor n/2 \rfloor$  such that  $Y_{k_1} = \dots = Y_{k_{m(j)}} = 1$  and  $Y_i = 0$  for all other indices. If we fix one such  $m(j)$ -tuple of indices, then this occurs exactly when the random variable

$$Y_{k_1} \cdot Y_{k_2} \cdots Y_{k_{m(j)}} \cdot \frac{\prod_{i=1}^{\lfloor n/2 \rfloor} (1 - Y_i)}{(1 - Y_{k_1})(1 - Y_{k_2}) \cdots (1 - Y_{k_{m(j)}})}$$

takes the value 1. Finally, adding over all the possible  $m(j)$ -tuples  $(k_1, \dots, k_{m(j)})$ , we obtain the random variable equal to  $Y_{\text{rk}=n-2j}(E(\mathbb{Q}))$ , as in the statement.

For the second part of the statement, we can calculate the expected value  $\mathbb{E}(Y_{\text{rk}=n-2j})$  as follows:

$$\begin{aligned} \mathbb{E}(Y_{\text{rk}=n-2j}) &= \mathbb{E} \left( \sum_{1 \leq k_1 < \dots < k_{m(j)} \leq \lfloor n/2 \rfloor} Y_{k_1} \cdot Y_{k_2} \cdots Y_{k_{m(j)}} \cdot \frac{\prod_{i=1}^{\lfloor n/2 \rfloor} (1 - Y_i)}{(1 - Y_{k_1})(1 - Y_{k_2}) \cdots (1 - Y_{k_{m(j)}})} \right) \\ &= \sum_{1 \leq k_1 < \dots < k_{m(j)} \leq \lfloor n/2 \rfloor} \mathbb{E} \left( Y_{k_1} \cdot Y_{k_2} \cdots Y_{k_{m(j)}} \cdot \frac{\prod_{i=1}^{\lfloor n/2 \rfloor} (1 - Y_i)}{(1 - Y_{k_1})(1 - Y_{k_2}) \cdots (1 - Y_{k_{m(j)}})} \right) \\ &= \sum_{1 \leq k_1 < \dots < k_{m(j)} \leq \lfloor n/2 \rfloor} \mathbb{E}_{m(j),j}^n(X) \\ &= \binom{\lfloor n/2 \rfloor}{j} \cdot \mathbb{E}_{m(j),j}^n(X), \end{aligned}$$

where we have used equicorrelation of random variables for the equality of the expected value of the product of any  $m(j)$  random variables, and parts (1) and (2) of Lemma 5.11. Now, let  $Y_i$  be independent (in particular,  $C_{s,t}^n(X) = 0$ ) or approximately uncorrelated (i.e.,  $C_{s,t}^n(X) \approx 0$ ). Then,

$$\begin{aligned} \mathbb{E}(Y_{\text{rk}=n-2j}) &\cong \sum_{1 \leq k_1 < \dots < k_{m(j)} \leq \lfloor n/2 \rfloor} \mathbb{E}(Y_{k_1}) \cdots \mathbb{E}(Y_{k_{m(j)}}) \cdot \frac{\prod_{i=1}^{\lfloor n/2 \rfloor} (1 - \mathbb{E}(Y_i))}{(1 - \mathbb{E}(Y_{k_1})) \cdots (1 - \mathbb{E}(Y_{k_{m(j)}}))} \\ &= \binom{\lfloor n/2 \rfloor}{m(j)} \rho_n(X)^{m(j)} (1 - \rho_n(X))^j = \binom{\lfloor n/2 \rfloor}{j} \rho_n(X)^{m(j)} (1 - \rho_n(X))^j, \end{aligned}$$

as claimed, where we have used the facts that (a) if  $\text{Cov}(Y, Y') = 0$  (or  $\approx 0$ ), then  $\mathbb{E}(YY') \cong \mathbb{E}(Y) \cdot \mathbb{E}(Y')$  and  $\mathbb{E}(1 - Y) = 1 - \mathbb{E}(Y)$ , and (b) that  $Y_i = Y_{\text{Hasse},n,X}(s_i)$  and therefore  $\mathbb{E}(Y_i) = \rho_n(X)$ .  $\square$

Let us simplify the formulas of Corollary 5.14 for  $n = 1, \dots, 5$ .

**Corollary 5.15.** *If we assume  $H_B$ , and  $\mathcal{S}_n^X$  is non-empty, then the probabilities*

$$p_n(r) = \text{Prob}(E \in \mathcal{R}_r^X \mid E \in \mathcal{S}_n^X) = (\mathcal{R}_r^X \cap \mathcal{S}_n^X) / \mathcal{S}_n^X$$

for  $n = 1, \dots, 5$  and  $0 \leq r \leq n$  are given by the formulas in Table 8.

*Proof.* When  $n = \text{selrank}(E(\mathbb{Q})) = 1$ , our assumption on the finiteness of  $\text{III}$  implies that the Mordell-Weil rank is 1. Thus,  $p_1(0) = 0$  and  $p_1(1) = 1$ . When  $n = 2$  or  $3$ , Theorem 5.9 says that there is a unique random variable  $Y_1$ , with mean  $\rho_n(X)$ , such that  $\text{rank}_{1,X} = Y_1$ . It follows that  $p_n(n) = \rho_n(X)$  and  $p_n(n - 2) = 1 - \rho_n(X)$ , and  $p_n(r) = 0$  for  $r \neq n - 2, n$ .



$p_n(r)$	2	3	4	5
$r = 0$	$1 - \rho_2(X)$	0	$(1 - \rho_4(X))^2 + C_{1,1}^4(X)$	0
1	0	$1 - \rho_3(X)$	0	$(1 - \rho_5(X))^2 + C_{1,1}^5(X)$
2	$\rho_2(X)$	0	$2\rho_4(X)(1 - \rho_4(X)) - 2C_{1,1}^4(X)$	0
3		$\rho_3(X)$	0	$2\rho_5(X)(1 - \rho_5(X)) - 2C_{1,1}^5(X)$
4			$\rho_4(X)^2 + C_{1,1}^4(X)$	0
5				$\rho_5(X)^2 + C_{1,1}^5(X)$

TABLE 8. Values of  $p_n(r) = \text{Prob}(\text{rank}(E(\mathbb{Q})) = r \mid E \in \mathcal{S}_n([X, X]))$  for  $n = 2, \dots, 5$  and  $0 \leq r \leq n$ . Note that  $p_1(0) = 0$  and  $p_1(1) = 1$ .

Finally, if  $n = 4, 5$ , then  $\text{rank}_{n,X} = Y_1 + Y_2$ , and Corollary 5.14 says that

$$Y_{\text{rk}=n} = Y_1 Y_2, \quad Y_{\text{rk}=n-2} = Y_1(1 - Y_2) + (1 - Y_1)Y_2, \quad Y_{\text{rk}=n-4} = (1 - Y_1)(1 - Y_2),$$

with expected value, respectively, given by

$$\begin{aligned} \mathbb{E}(Y_{\text{rk}=n}) &= \mathbb{E}(Y_1)\mathbb{E}(Y_2) + \text{Cov}(Y_1, Y_2) = \rho_n(X)^2 + C_{1,1}^n(X), \\ \mathbb{E}(Y_{\text{rk}=n-2}) &= \mathbb{E}(Y_1)(1 - \mathbb{E}(Y_2)) - \text{Cov}(Y_1, Y_2) + (1 - \mathbb{E}(Y_1))\mathbb{E}(Y_2) - \text{Cov}(Y_1, Y_2) \\ &= 2\rho_n(X)(1 - \rho_n(X)) - 2C_{1,1}^n(X), \\ \mathbb{E}(Y_{\text{rk}=n-4}) &= (1 - \mathbb{E}(Y_1))(1 - \mathbb{E}(Y_2)) + \text{Cov}(Y_1, Y_2) = (1 - \rho_n(X))^2 + C_{1,1}^n(X), \end{aligned}$$

where we have used the equality  $\mathbb{E}(Y_1 Y_2) = \mathbb{E}(Y_1)\mathbb{E}(Y_2) + \text{Cov}(Y_1, Y_2)$  and the fact that the covariance satisfies  $\text{Cov}(a + bY_1, c + dY_2) = bd \text{Cov}(Y_1, Y_2)$  for constants  $a, b, c, d$ .  $\square$

**Remark 5.16.** The formulas for the expected value of  $Y_{\text{rk}=n-2j}$  for  $n \geq 6$ , unfortunately, cannot be written just in terms of  $\mathbb{E}(Y_i)$  and  $C_{1,1}^n(X) = \text{Cov}(Y_i, Y_j)$  for  $i \neq j$ . One needs to know other higher moments of the random variables  $Y_i$ . For instance, let  $n = 6$ . Then,

$$\begin{aligned} \mathbb{E}(Y_{\text{rk}=6}) &= \mathbb{E}(Y_1 Y_2 Y_3) = \mathbb{E}(Y_1 Y_2)\mathbb{E}(Y_3) + \text{Cov}(Y_1 Y_2, Y_3) \\ &= \mathbb{E}(Y_1)\mathbb{E}(Y_2)\mathbb{E}(Y_3) + \text{Cov}(Y_1, Y_2)\mathbb{E}(Y_3) + \text{Cov}(Y_1 Y_2, Y_3) \\ &= \rho_6(X)^3 + C_{1,1}^6(X)\rho_6(X) + C_{2,1}^6(X). \end{aligned}$$

The formulae for  $\mathbb{E}(Y_{\text{rk}=6-2j})$  can be written in terms of the functions  $\rho_6(X)$ ,  $C_{1,1}^6$ , and  $C_{2,1}^6(X)$ . For example,

$$\begin{aligned} \mathbb{E}(Y_{\text{rk}=4}) &= \mathbb{E}(Y_1 Y_2(1 - Y_3)) + \mathbb{E}(Y_1(1 - Y_2)Y_3) + \mathbb{E}((1 - Y_1)Y_2 Y_3) = 3 \cdot \mathbb{E}(Y_1 Y_2(1 - Y_3)) \\ &= 3(\mathbb{E}(Y_1 Y_2)\mathbb{E}(1 - Y_3) + \text{Cov}(Y_1 Y_2, 1 - Y_3)) \\ &= 3(\mathbb{E}(Y_1 Y_2)(1 - \mathbb{E}(Y_3)) - \text{Cov}(Y_1 Y_2, Y_3)) \\ &= 3((\mathbb{E}(Y_1)\mathbb{E}(Y_2) + \text{Cov}(Y_1, Y_2))(1 - \mathbb{E}(Y_3)) - \text{Cov}(Y_1 Y_2, Y_3)) \\ &= 3(\rho_6(X)^2(1 - \rho_6(X)) + C_{1,1}^6(X)(1 - \rho_6(X)) - C_{2,1}^6(X)). \end{aligned}$$

In order to estimate the values of  $\rho_n(X)$ , we define the following moving ratio measuring the failure of the Hasse principle for 2-Selmer elements coming from elliptic curves of Selmer rank  $n$  and up to height  $X$ .

**Definition 5.17.** For each  $n \geq 2$ , and  $N \geq 0$ , we define the average failure of the Hasse principle for 2-Selmer elements up in the height interval  $(X, X + N]$  by

$$\begin{aligned} \rho_n(X, N) &= \frac{\sum_{E \in \mathcal{S}_n((X, X+N])} (\text{rank}(E(\mathbb{Q})) - (n \bmod 2))}{\sum_{E \in \mathcal{S}_n((X, X+N])} (\text{selrank}(E(\mathbb{Q})) - (n \bmod 2))} \\ &= \frac{\sum_{E \in \mathcal{S}_n((X, X+N])} (\text{rank}(E(\mathbb{Q})) - (n \bmod 2))}{(n - (n \bmod 2)) \cdot \pi_{\mathcal{S}_n}((X, X + N])}. \end{aligned}$$

**Corollary 5.18.** If we assume  $H_B$ , and  $X > N^2 \geq 0$ , then the expected value of  $\rho_n(X, N)$  is given by  $\rho_n(X) + O(X^{-1/3})$  on average, with a standard error

$$\begin{aligned} &\approx \sqrt{\frac{6X^{1/6} \lfloor n/2 \rfloor (\rho_n(X)(1 - \rho_n(X)) + (\lfloor n/2 \rfloor - 1)C_{1,1}^n(X))}{5\kappa N \theta_n(X)}} + O\left(\frac{1}{NX^{1/6}}\right) \\ &\approx \sqrt{\frac{6X^{1/6}(1 + C_n X^{-e_n}) \lfloor n/2 \rfloor (\rho_n(X)(1 - \rho_n(X)) + (\lfloor n/2 \rfloor - 1)C_{1,1}^n(X))}{5\kappa s_n N}} + O\left(\frac{1}{NX^{1/6}}\right) \end{aligned}$$

where  $C_{1,1}^n(X) = 0$  for  $n = 2, 3$ , and the last approximation assumes  $H_A$  and Conjecture 4.5.

*Proof.* By Corollary 5.10, the expected value of

$$\sum_{E \in \mathcal{S}_n((X, X+N])} \frac{\text{rank}(E(\mathbb{Q})) - (n \bmod 2)}{2}$$

is given by

$$\sum_{E \in \mathcal{S}_n((X, X+N])} \lfloor n/2 \rfloor \rho_n(\text{ht}(E)) = \lfloor n/2 \rfloor \sum_{H=X+1}^{X+N} \sum_{E \in \mathcal{S}_n^H} \rho_n(H) = \lfloor n/2 \rfloor \sum_{H=X+1}^{X+N} \pi_{\mathcal{S}_n}([H, H]) \cdot \rho_n(H).$$

By Proposition 4.9, the expected value of  $\pi_{\mathcal{S}_n}([H, H]) \approx \int_{H-1}^H \frac{5\kappa \theta_n(T)}{6T^{1/6}} dT + O\left(\frac{1}{H^{1/2}}\right)$  on average, and since the limit of  $\rho_n(H) = 0$  by  $H_B$ , then Lemma 3.6 says that

$$\lfloor n/2 \rfloor \sum_{H=X+1}^{X+N} \pi_{\mathcal{S}_n}([H, H]) \cdot \rho_n(H) \approx \frac{5\kappa \lfloor n/2 \rfloor}{6} \int_X^{X+N} \frac{\theta_n(T) \rho_n(T)}{T^{1/6}} dT + O\left(\frac{N \rho_n(H)}{H^{1/2}}\right)$$

on average, which in turn (as in Corollary 3.4) says that, for  $X > N^2 \geq 0$ , we have

$$\mathbb{E} \left( \sum_{E \in \mathcal{S}_n((X, X+N])} \frac{\text{rank}(E(\mathbb{Q})) - (n \bmod 2)}{2} \right) \approx \frac{5\kappa \lfloor n/2 \rfloor N}{6} \frac{\theta_n(X) \rho_n(X)}{X^{1/6}} + O\left(\frac{N \rho_n(X)}{X^{1/2}}\right).$$

By Proposition 4.9 we have that  $\mathbb{E}(\pi_{\mathcal{S}_n}((X, X + N])) \approx 5\kappa N / (6X^{1/6}) + O(N/X^{1/2})$ , and therefore, the expected value of

$$\rho_n(X, N) = \frac{2}{(n - (n \bmod 2)) \pi_{\mathcal{S}_n}((X, X+N])} \sum_{E \in \mathcal{S}_n((X, X+N])} \frac{\text{rank}(E(\mathbb{Q})) - (n \bmod 2)}{2}$$

is given by

$$\rho_n(X) + O\left(\frac{X^{1/6}\rho_n(X)}{X^{1/2}}\right) = \rho_n(X) + O\left(\frac{\rho_n(X)}{X^{1/3}}\right),$$

on average, where we have used the simple fact that  $2\lfloor n/2 \rfloor = n - (n \bmod 2)$ . Since  $\lim_{X \rightarrow \infty} \rho_n(X) = 0$ , we can simplify the error term to  $O(X^{-1/3})$ .

The standard error can be deduced from the formula of Corollary 5.10 for  $\pi_{\mathcal{S}_n}((X, X+N])$  curves, the number of Selmer curves on average from Prop. 4.9, and Lemma 3.6, and it is given on average by

$$\begin{aligned} &= \frac{1}{\pi_{\mathcal{S}_n}((X, X+N])} \sqrt{\lfloor n/2 \rfloor \sum_{H=X+1}^{X+N} \pi_{\mathcal{S}_n}([H, H]) \cdot (\rho_n(H)(1 - \rho_n(H)) + (\lfloor n/2 \rfloor - 1)C_{1,1}^n(H))} \\ &\approx \frac{1}{\pi_{\mathcal{S}_n}((X, X+N])} \sqrt{\frac{5\kappa N \theta_n(X)}{6X^{1/6}} \cdot (\rho_n(X)(1 - \rho_n(X)) + (\lfloor n/2 \rfloor - 1)C_{1,1}^n(X)) + O\left(\frac{N}{X^{1/2}}\right)} \\ &\approx \sqrt{\frac{6X^{1/6}\lfloor n/2 \rfloor(\rho_n(X)(1 - \rho_n(X)) + (\lfloor n/2 \rfloor - 1)C_{1,1}^n(X))}{5\kappa N \theta_n(X) + O(NX^{-1/3})}} + O\left(\frac{1}{NX^{1/3}}\right) \\ &\approx \sqrt{\frac{6X^{1/6}\lfloor n/2 \rfloor(\rho_n(X)(1 - \rho_n(X)) + (\lfloor n/2 \rfloor - 1)C_{1,1}^n(X))}{5\kappa N \theta_n(X)}} + O\left(\frac{1}{NX^{1/6}}\right) \\ &\approx \sqrt{\frac{6X^{1/6}(1 + C_n X^{-e_n})\lfloor n/2 \rfloor(\rho_n(X)(1 - \rho_n(X)) + (\lfloor n/2 \rfloor - 1)C_{1,1}^n(X))}{5\kappa s_n N}} + O\left(\frac{1}{NX^{1/6}}\right) \end{aligned}$$

as claimed, where in the approximations we assumed  $H_A$  and we have used the results of Proposition 4.9, part (2).  $\square$

In order to test Hypothesis B, we have used the BHKSSW data to estimate probability function  $\rho_n(X)$  using the moving ratios  $\rho_n(X, N)$  of Corollary 5.18. We have plotted values of  $\rho_n(X, 0.25 \cdot 10^9)$  for  $n = 2, \dots, 5$  using the BHKSSW database, and the graphs can be found in Figure 9.

In Table 9 we record the last values of  $\rho_n(X, 0.25 \cdot 10^9)$  that appear in the graphs (which correspond to  $X \approx 2.675 \cdot 10^{10}$ ). We also record the values of  $\pi_{\mathcal{S}_n}$  in  $[2.675 \cdot 10^{10}, 2.7 \cdot 10^{10}]$ . The total number of elliptic curves in the same interval is 1,828,235.

$n$	2	3	4	5
$\pi_{\mathcal{S}_n}([2.675 \cdot 10^{10}, 2.7 \cdot 10^{10}])$	476,579	104,922	7945	152
$\rho_n(2.675 \cdot 10^{10}, 0.25 \cdot 10^9)$	0.63989181	0.45496654	0.63857772	0.63486842

TABLE 9. The number of curves of Selmer rank  $2 \leq n \leq 5$ , and the values of  $\rho_n(X, N)$  in the interval  $[2.675 \cdot 10^{10}, 2.7 \cdot 10^{10}]$ .

Finally, we have found (using SageMath) best-fit models for the data of  $\rho_n(X, N)$  of the form

$$\rho_n(X, N) \approx \frac{D_n}{X f_n}.$$

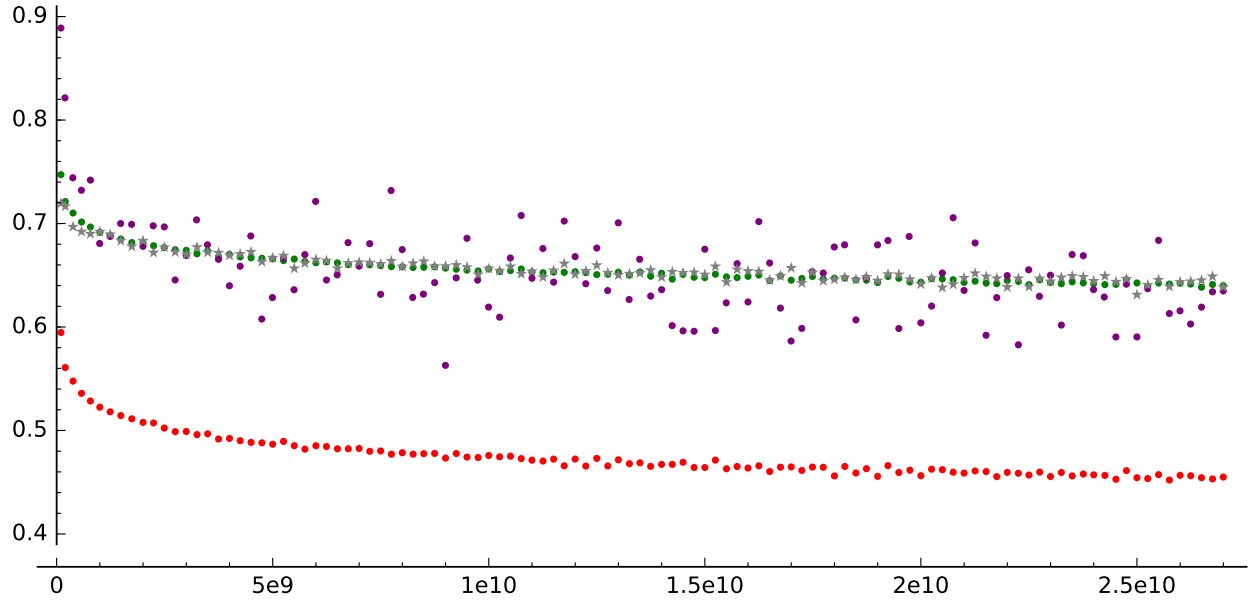


FIGURE 9. Graphs of the moving ratios  $\rho_n(X, 0.25 \cdot 10^9)$  for  $n = 2$  (green), 3 (red), 4 (gray stars), 5 (purple).

and we provide the values of  $D_n$  and  $f_n$  in Table 10. We have compared the models with the data in Figure 10.

$n$	2	3	4	5
$D_n$	1.12465347	1.30937016	1.07928016	1.79161787
$f_n$	0.02344245	0.04412662	0.02158211	0.04383626

TABLE 10. The coefficients of the best-fit models  $\rho_n(X, N) \approx D_n/X^{f_n}$ .

**Conjecture 5.19.** *Hypothesis  $H_B$  holds and, for every  $n \geq 2$ , there are constants  $D_n$  and  $f_n$  such that  $\rho_n(X) \approx \frac{D_n}{X^{f_n}}$ . Moreover, for  $n = 2, \dots, 5$  the values of  $D_n$  and  $f_n$  are approximately as given in Table 10.*

**Remark 5.20.** Before we can discuss the error in the approximation  $\rho_n(X, N) \approx \rho_n(X)$  we need to estimate the covariance functions  $C_{s,t}^n(X)$ . This can be done via the formulas for the expected value of  $Y_{\text{rk}=n-2j}(E(\mathbb{Q}))$  given by Corollary 5.14 and, for  $n = 1, 2, 3, 4, 5$ , the simplified formulas given by Corollary 5.15. The first thing to note is that for  $n = 1, 2, 3$ , we have  $C_{s,t}^n(X) = 0$  for all possible values of  $s, t$  since there is either none ( $n = 1$ ) or only one random variable  $Y_1$  that intervenes ( $n = 2, 3$ ). For  $n = 4$  and 5 there are two random variables  $Y_1$  and  $Y_2$  and

$$C_{1,1}^n(X) = \mathbb{E}(Y_{\text{rk}=n}(E(\mathbb{Q}))) - \mathbb{E}(Y_1)\mathbb{E}(Y_2) = \mathbb{E}(Y_{\text{rk}=n}(E(\mathbb{Q}))) - \rho_n(X)^2.$$

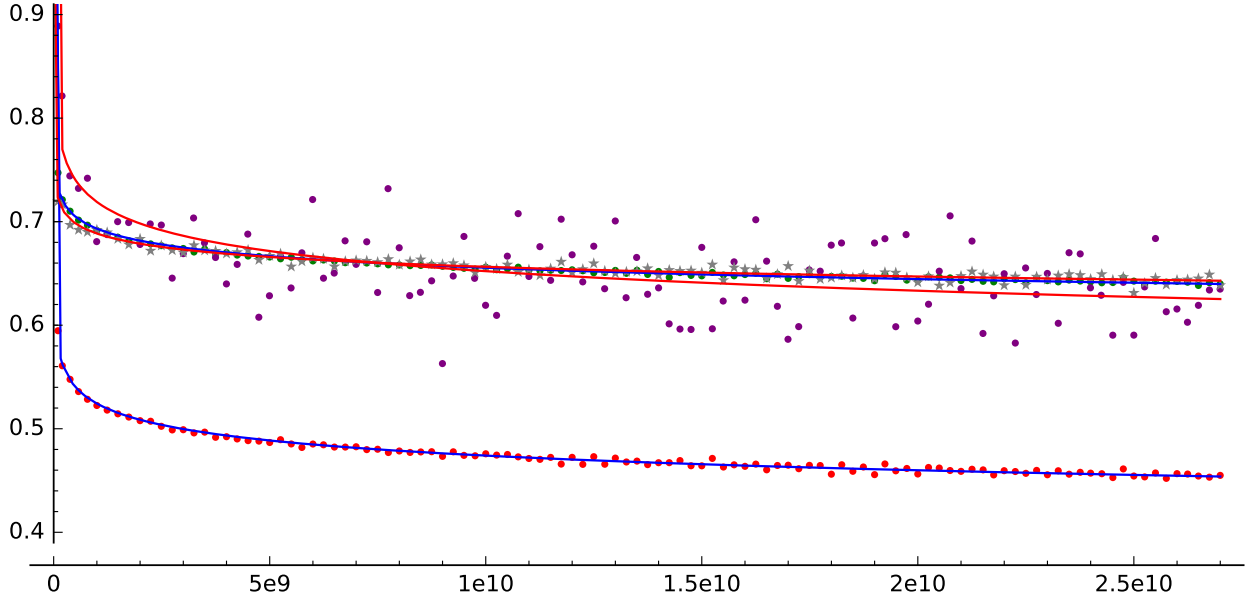


FIGURE 10. Graphs of the moving ratios  $\rho_n(X, 0.025 \cdot 10^9)$  for  $n = 2$  (green), 3 (red), 4 (gray stars), 5 (purple), and the corresponding models of the form  $D_n/X^{f_n}$  (in blue for  $n = 2, 3$  and red for  $n = 4, 5$ ).

In Figures 11 and 12 we have plotted approximate covariance values of  $C_{1,1}^4(X)$  and  $C_{1,1}^5(X)$ , respectively, using sample height intervals  $[X, X + 0.25 \cdot 10^9]$ , together with the best linear fits for the data which are given by

$$-0.02677186 + (4.06113344 \cdot 10^{-14})x \quad \text{and} \quad -0.01328180 + (1.28980002 \cdot 10^{-12})x,$$

respectively. In particular, we observe that  $|C_{1,1}^4(X) - (-0.025)| \lesssim 0.015$  and  $|C_{1,1}^5(X) - 0| \lesssim 0.1$ . Thus, below, we will approximate  $C_{1,1}^4(X) \approx -0.025$  and  $C_{1,1}^5(X) \approx 0$ .

**Remark 5.21.** Let us assume Conjecture 5.19, and let us use Corollary 5.18 to estimate the error in the approximation  $\rho_n(X) \approx \rho_n(X, N)$ . The error should be given by the expression

$$\text{err}_{1,n}(X, N) = \sqrt{\frac{[n/2](\rho_n(X)(1 - \rho_n(X)) + ([n/2] - 1)C_{1,1}^n(X))}{\pi_{\mathcal{S}_n}((X, X + N])}}$$

or by the expression

$$\text{err}_{2,n}(X, N) = \sqrt{\frac{6X^{1/6}(1 + C_n X^{-e_n})[n/2](\rho_n(X)(1 - \rho_n(X)) + ([n/2] - 1)C_{1,1}^n(X))}{5\kappa s_n N}},$$

if we assume  $H_A$  and Conjecture 4.5 also. Using our calculations of Remark 5.20, we will take  $C_{1,1}^n(X) = 0$  for  $n = 2, 3$ , and  $C_{1,1}^4(X) = -0.025$ , and  $C_{1,1}^5(X) = 0$ . In Table 11 we include the values of:  $\rho_n(X, N)$ , our model of  $\rho_n(X)$ , the error of the model  $|\rho_n(X, N) - \rho_n(X)|$ , and the predicted standard errors  $\text{err}_{i,n}(X, N)$ , for  $i = 1, 2$ , and  $X = 2.675 \cdot 10^{10}$ , with  $N = 0.25 \cdot 10^9$ .

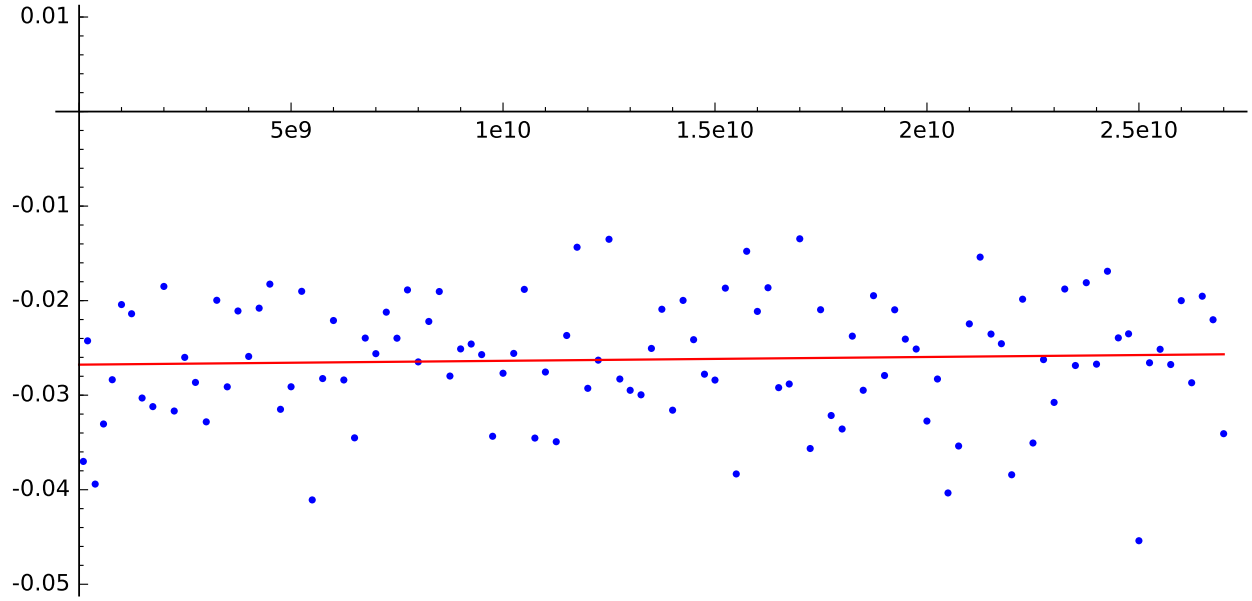


FIGURE 11. Approximate values of  $C_{1,1}^4(X)$  using sample height intervals  $[X, X + 0.25 \cdot 10^9]$  to estimate  $\mathbb{E}(Y_{\text{rk}=4}(E(\mathbb{Q}))) - \rho_4(X)^2$ . The best-fit line is given by  $-0.02677186 + (4.06113344 \cdot 10^{-14})x$ .

$n$	2	3	4	5
$\pi_{S_n}([2.675 \cdot 10^{10}, 2.7 \cdot 10^{10}])$	476,579	104,922	7945	152
$\rho_n(2.675 \cdot 10^{10}, 0.25 \cdot 10^9)$	0.63989181	0.45496654	0.63857772	0.63486842
$\rho_n(2.675 \cdot 10^{10})$	0.63996477	0.45404630	0.64309203	0.62550968
$ \text{Error}  =  \rho_n - \rho_n $	0.00007296	0.00092023	0.00451431	0.00935873
$\text{err}_{1,n}(2.675 \cdot 10^{10}, 0.25 \cdot 10^9)$	0.00069531	0.00153707	0.00717531	0.05551757
$\text{err}_{2,n}(2.675 \cdot 10^{10}, 0.25 \cdot 10^9)$	0.00069208	0.00152440	0.00700827	0.05462609

TABLE 11. Values of:  $\rho_n(X, N)$ , our model of  $\rho_n(X)$ , the error  $|\rho_n(X, N) - \rho_n(X)|$ , and the two predicted standard errors  $\text{err}_{i,n}(X, N)$ , for  $i = 1, 2$ , and  $X = 2.675 \cdot 10^{10}$ ,  $N = 0.25 \cdot 10^9$ .

**Remark 5.22.** As we can see from the errors in Table 11, there is certainly insufficient data for  $n = 5$ , so our models of  $\rho_5(X)$  are not as accurate as we would wish.

**Remark 5.23.** As we have mentioned earlier in Remark 4.7 the BHKSSW database ([1]) also includes small databases of random samples of elliptic curves at larger heights. In order to test  $H_B$  and Conjecture 5.19, we have calculated the average Hasse ratio for the curves in  $\mathcal{E}_k$  (with notation

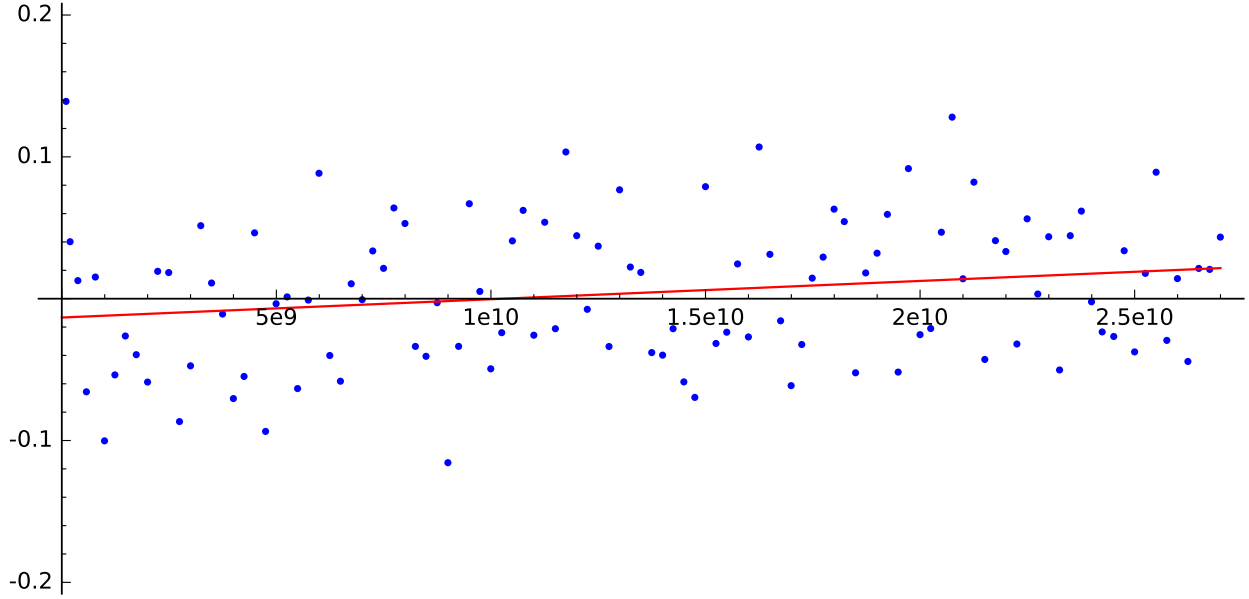


FIGURE 12. Approximate values of  $C_{1,1}^5(X)$  using sample height intervals  $[X, X + 0.25 \cdot 10^9]$  to estimate  $\mathbb{E}(Y_{\text{rk}=5}(E(\mathbb{Q}))) - \rho_5(X)^2$ . The best-fit line is given by  $-0.01328180 + (1.28980002 \cdot 10^{-12})x$ .

as in Remark 4.7), and have plotted the ratios together with our models for  $\rho_n(X)$ , in Figure 13 (note: the  $x$ -axis is in logarithmic scale). We have also computed the predicted errors (a calculation similar to that carried out in Table 11) and the predictions seem to match the data in large heights, as well.

**Remark 5.24.** It would be interesting to compute the ratio  $\rho_n(X)$  in families of quadratic twists. However, such families are very “thin” in the family of all elliptic curves, and the convergence of the Hasse ratios to  $\rho_n(X)$  would be unreliable. In order to provide some data in this direction, we have calculated the Selmer rank and Mordell-Weil rank in a family of twists (quadratic and quartic) of  $y^2 = x^3 + x$ . More precisely, we consider the curves  $E_A : y^2 = x^3 + Ax$ , with fourth-power-free  $1 \leq A \leq 10^6$  (curves up to height  $4 \cdot 10^{18}$ ). Then, we have calculated the moving ratios  $\rho_n$  in slices of 10,000 curves, and graphed them against the models of Conjecture 5.19. See Figure 14. Note, however, that we do not expect the exact same behavior in this family, since  $j(E_A/\mathbb{Q}) = 1728$  is fixed, and therefore it is a family of twists (quadratic and quartic). It is likely that if  $H_B$  holds, then a similar condition is true for  $j = 1728$  up to a constant. That is, we may have  $\rho_{n,1728}(X) \approx C_{1728} \cdot \rho_n(X)$ , where  $C_{1728}$  is a fixed constant. At any rate, the family of curves with  $j = 1728$  is very sparse within the family of all curves, and the data only indicates some consistency with our expectations.

**Example 5.25.** Theorem 5.9, assuming  $H_B$ , provides the expected value and variance for the rank of an elliptic curve  $E/\mathbb{Q}$  of Selmer rank  $n$  and height  $X$ . More precisely, in Corollary 5.14 and 5.15, we give formulas for the probabilities for each rank. Now that we have models for  $\rho_n(X)$  and  $C_{1,1}^n(X)$  (as in Remark 5.20), we can look at the distribution of ranks in intervals. Let us consider,

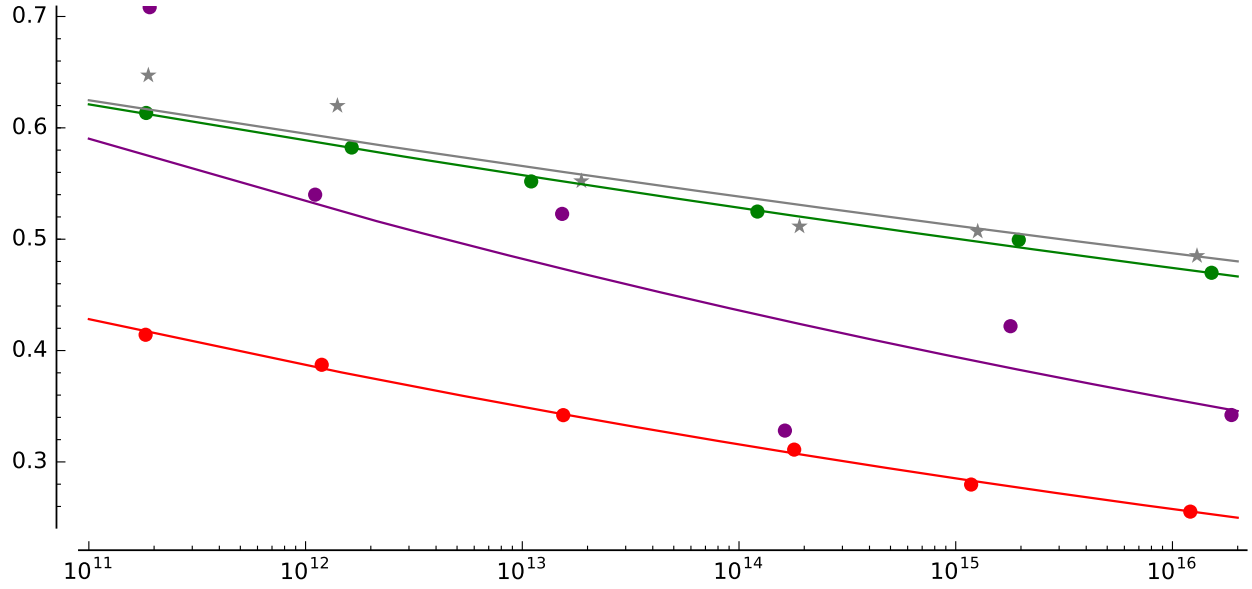


FIGURE 13. Graphs of the moving ratios  $\rho_n(X, N)$  for the curves of large height in the database BHKSSW, for  $n = 2$  (green), 3 (red), 4 (gray stars), 5 (purple), and the corresponding models of the form  $D_n/X^{f_n}$  (in blue for  $n = 2, 3$  and red for  $n = 4, 5$ ). The  $x$ -axis is in logarithmic scale.

for instance, the curves  $\mathcal{E}(I)$  in the height interval  $I = [20 \cdot 10^9, 20.25 \cdot 10^9]$  in the BHKSSW database. For each  $n = 2, 3, 4, 5$  we have created histograms using the number of curves of Selmer rank  $n$  and Mordell-Weil rank  $0 \leq n$  (in blue bars), and also created histograms with the number of M-W ranks that we would expect from Corollary 5.15 (in green bars). The resulting histograms can be found in Figure 15 (together with the graph of the normal distribution that would approximate the binomial  $B(\lfloor n/2 \rfloor, \rho_n(X))$ ). We have also included the raw data of ranks observed and ranks predicted in Table 12.

$n$	$\pi_{\mathcal{S}_n}([2 \cdot 10^{10}, 2.025 \cdot 10^{10}])$	M-W ranks observed in $\mathcal{S}_n$	M-W ranks predicted
2	509,845	[180128, 0, 329717, 0, 0, 0]	[181246.58, 0, 328598.41, 0, 0, 0]
3	111,926	[0, 60149, 0, 51777, 0, 0]	[0, 60455.09, 0, 51470.90, 0, 0]
4	8399	[803, 0, 4321, 0, 3275, 0]	[836.68, 0, 4256.52, 0, 3305.78, 0]
5	158	[0, 22, 0, 76, 0, 60]	[0, 21.24, 0, 73.38, 0, 63.36]

TABLE 12. Mordell-Weil ranks observed in the interval height interval  $[2 \cdot 10^{10}, 2.025 \cdot 10^{10}]$  and the ranks predicted by the distribution of Theorem 5.9.



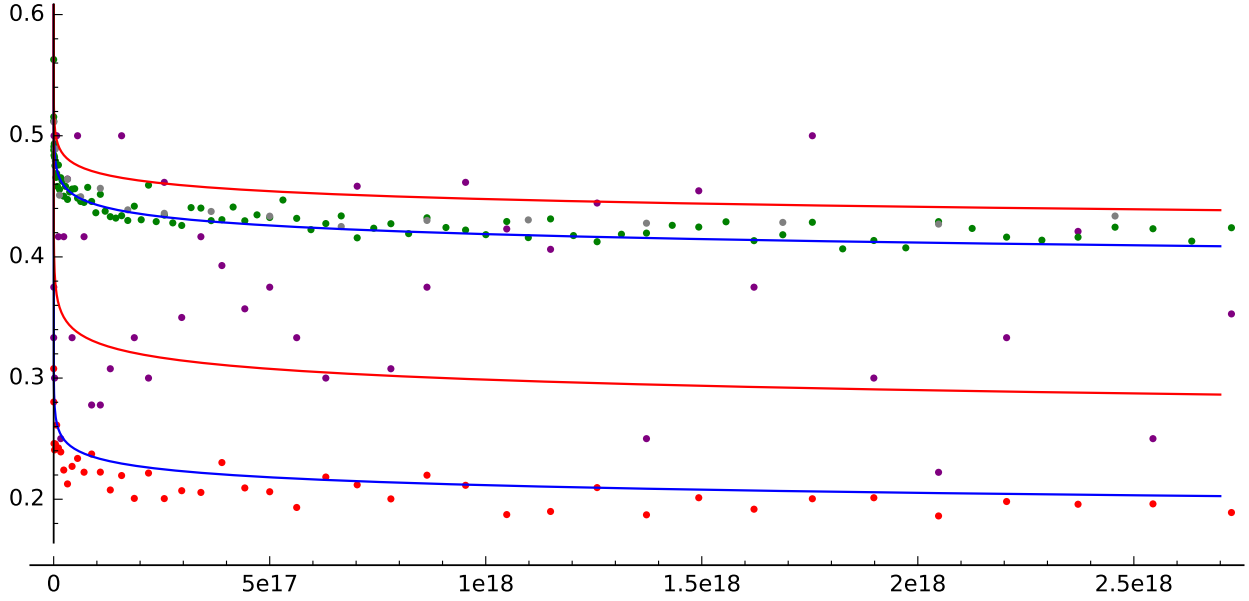


FIGURE 14. Graphs of the moving ratios  $\rho_n$  in the family  $y^2 = x^3 + Ax$  for  $n = 2$  (green), 3 (red), 4 (gray stars), 5 (purple), compared to the models of the form  $D_n/X^{f_n}$  (in blue for  $n = 2, 3$  and red for  $n = 4, 5$ ).

## 6. PREDICTING THE NUMBER OF CURVES WITH A GIVEN RANK UP TO HEIGHT $X$

Let  $X, r \geq 0$  be fixed. We denote the set of elliptic curves of height  $\leq X$  and Mordell-Weil rank  $r$  by

$$\mathcal{R}_r(X) = \{E \in \mathcal{E}(X) : \text{rank}(E(\mathbb{Q})) = r\},$$

and we write  $\pi_{\mathcal{R}_r}(X) = \#\mathcal{R}_r(X)$ . We refer the reader to Sections 3.3 and 3.4 of [21] for a summary of conjectures about  $\pi_{\mathcal{R}_r}(X)$ , but we point out two in particular:

- Watkins ([27]; see also [2] for an expository paper) has conjectured that there is a constant  $c$  such that

$$\sum_{k=1}^{\infty} \pi_{\mathcal{R}_{2k}}(X) = (c + o(1))X^{19/24}(\log X)^{3/8}.$$

- Park, Poonen, Voight, and Wood ([21]) have developed a heuristic that predicts:
  - (1) All but finitely many elliptic curves satisfy  $\text{rank}(E(\mathbb{Q})) \leq 21$ .
  - (2) For  $1 \leq r \leq 20$ , we have  $\sum_{k=r}^{\infty} \pi_{\mathcal{R}_k}(X) = X^{(21-r)/24+o(1)}$ .
  - (3)  $\sum_{k=21}^{\infty} \pi_{\mathcal{R}_k}(X) \leq X^{o(1)}$ .

In this section, we shall assume  $H_A$  and  $H_B$ , and derive the expected value of  $\pi_{\mathcal{R}_r}(X)$  that follows from the probability distributions we have studied in previous sections. We shall study  $\pi_{\mathcal{R}_r}(X)$  as the sum of the contributions of Mordell-Weil rank  $r$  coming from each Selmer rank  $n = r + 2j$ . That

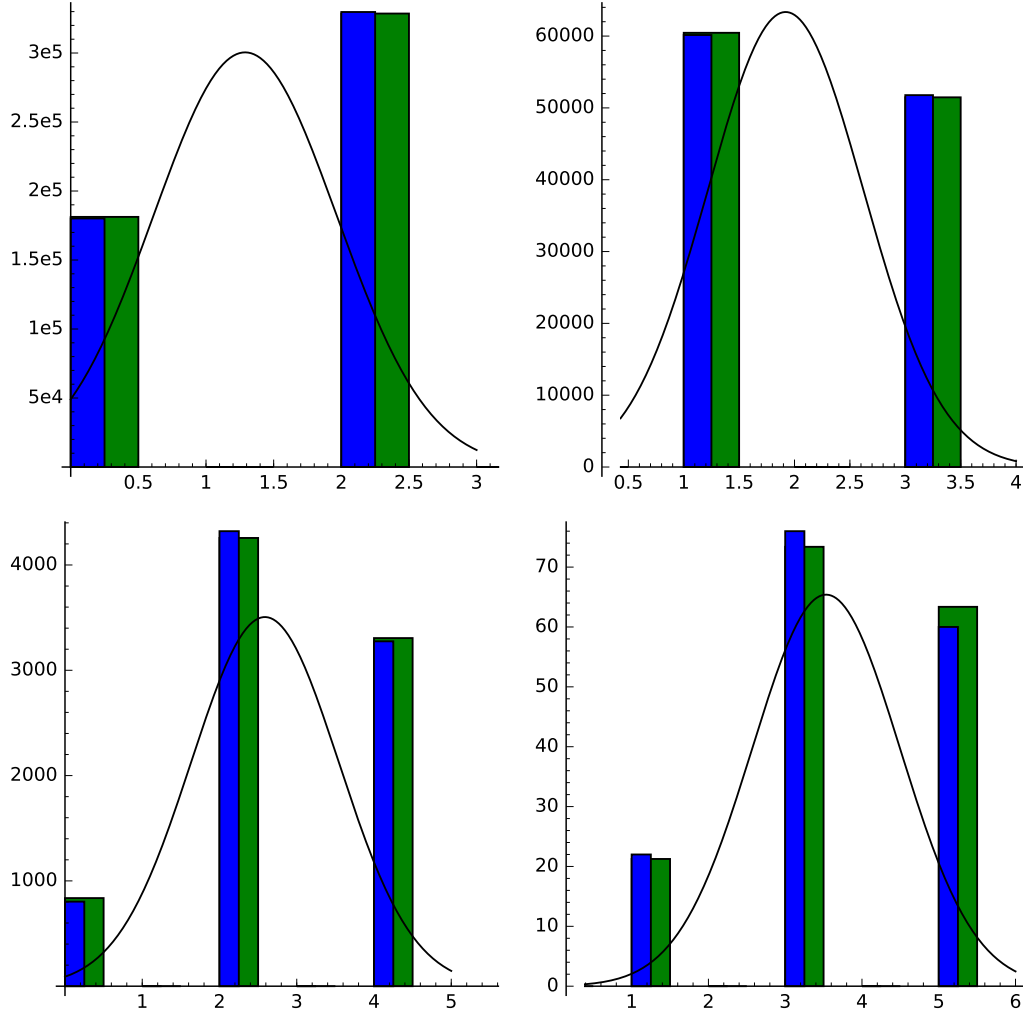


FIGURE 15. Histogram (in blue) of the distribution of Mordell-Weil ranks among elliptic curves in  $\mathcal{E}([2 \cdot 10^{10}, 2.025 \cdot 10^{10}])$  by Selmer rank  $n = 2, 3, 4, 5$ , and compared to the histogram (in green) of the M-W ranks that we would expect from Theorem 5.9. The graph is that of the normal distribution that best approximates the binomial.

is, we shall approximate  $\pi_{\mathcal{R}_r}(X)$  by approximating each term in the infinite sum

$$\pi_{\mathcal{R}_r}(X) = \sum_{j=0}^{\infty} \pi_{\mathcal{R}_r \cap \mathcal{S}_{r+2j}}(X).$$

Thus, for fixed  $r \geq 0$ , we first give the expected value of  $\pi_{\mathcal{R}_r \cap \mathcal{S}_{r+2j}}(X)$  for each  $j \geq 0$ .

**Theorem 6.1.** *Let  $X, r \geq 0, j \geq 0$  be fixed, such that  $n(j) = r + 2j \geq 2$ . If we assume  $H_A$  and  $H_B$ , then the expected value of  $\pi_{\mathcal{R}_r \cap \mathcal{S}_{n(j)}}(X)$  is given by*

$$\mathbb{E} \left( \pi_{\mathcal{R}_r \cap \mathcal{S}_{n(j)}}(X) \right) = \frac{5\kappa}{6} \binom{\lfloor \frac{r}{2} \rfloor + j}{j} \int_0^X \frac{\theta_{n(j)}(H)}{H^{1/6}} \cdot \mathbb{E}_{\lfloor \frac{r}{2} \rfloor, j}^{n(j)}(H) dH + \theta_{n(j)}(X) \cdot O \left( X^{1/2} \right),$$

where  $\mathbb{E}_{\lfloor \frac{r}{2} \rfloor, j}^{n(j)}(H)$  is the expected value defined in Remark 5.12. Further, if we assume Conjecture 4.5, then

$$\pi_{\mathcal{R}_r \cap \mathcal{S}_{n(j)}}(X) \approx \frac{5\kappa}{6} \binom{\lfloor \frac{r}{2} \rfloor + j}{j} \int_0^X \frac{s_{n(j)} \cdot \mathbb{E}_{\lfloor \frac{r}{2} \rfloor, j}^{n(j)}(H)}{(1 + C_{n(j)} H^{-e_{n(j)}}) \cdot H^{1/6}} dH + \theta_{n(j)}(X) \cdot O \left( X^{1/2} \right).$$

*Proof.* Let us write  $n(j) = r + 2j$ . Thus,  $\lfloor \frac{n(j)}{2} \rfloor = \lfloor \frac{r}{2} \rfloor + j$ . We compute the expected value of  $\pi_{\mathcal{R}_r}(X)$  as follows:

$$\begin{aligned} \mathbb{E} \left( \pi_{\mathcal{R}_r \cap \mathcal{S}_{n(j)}}(X) \right) &= \mathbb{E} \left( \#\{E \in \mathcal{S}_{n(j)}(X) : \text{rank}(E(\mathbb{Q})) = r\} \right) \\ &= \sum_{T=1}^X \mathbb{E} \left( \#\{E \in \mathcal{S}_{n(j)}([T, T]) : \text{rank}(E(\mathbb{Q})) = r\} \right) \\ &= \sum_{T=1}^X \mathbb{E} \left( \pi_{\mathcal{S}_{n(j)}}([T, T]) \right) \cdot \text{Prob}(\text{rank}(E(\mathbb{Q})) = r \mid E \in \mathcal{S}_{r+2j}([T, T])), \end{aligned}$$

by the basic properties of the expected value. If we assume  $H_A$  and  $H_B$  and use Corollary 4.4 for the value of  $\pi_{\mathcal{S}_{n(j)}}([H, H])$  (on average) and Corollary 5.14 for the probability of rank  $r$  in  $\mathcal{S}_{n(j)}([X, X])$ , we obtain the following formula:

$$\begin{aligned} \mathbb{E} \left( \pi_{\mathcal{R}_r \cap \mathcal{S}_{n(j)}}(X) \right) &= \sum_{T=0}^{X-1} \left( \frac{5\kappa}{6} \int_T^{T+1} \frac{\theta_{n(j)}(H)}{H^{1/6}} \cdot \binom{\lfloor \frac{n(j)}{2} \rfloor}{j} \cdot \mathbb{E}_{\lfloor \frac{r}{2} \rfloor, j}^{n(j)}(H) dH + \theta_{n(j)}(T) \cdot O \left( \frac{1}{T^{1/2}} \right) \right) \\ &= \frac{5\kappa}{6} \binom{\lfloor \frac{r}{2} \rfloor + j}{j} \int_0^X \frac{\theta_{n(j)}(H)}{H^{1/6}} \cdot \mathbb{E}_{\lfloor \frac{r}{2} \rfloor, j}^{n(j)}(H) dH + \theta_{n(j)}(X) \cdot O \left( X^{1/2} \right). \end{aligned}$$

If we assume Conjecture 4.5, then

$$\pi_{\mathcal{R}_r \cap \mathcal{S}_{n(j)}}(X) \approx \frac{5\kappa}{6} \binom{\lfloor \frac{r}{2} \rfloor + j}{j} \int_0^X \frac{s_{n(j)} \cdot \mathbb{E}_{\lfloor \frac{r}{2} \rfloor, j}^{n(j)}(H)}{(1 + C_{n(j)} H^{-e_{n(j)}}) \cdot H^{1/6}} dH + \theta_{n(j)}(X) \cdot O \left( X^{1/2} \right),$$

as claimed.  $\square$

If we now use the formula  $\pi_{\mathcal{R}_r}(X) = \sum_{j=0}^{\infty} \pi_{\mathcal{R}_r \cap \mathcal{S}_{r+2j}}(X)$  and the fact that  $\sum_{n=0}^{\infty} \theta_n(X) = 1$  (from Corollary 4.4), we obtain the following result.

**Corollary 6.2.** *Let  $X, r \geq 0$  be fixed. If we assume  $H_A$  and  $H_B$ , then the expected value of  $\pi_{\mathcal{R}_r}(X)$  is given by the formula*

$$\mathbb{E}(\pi_{\mathcal{R}_r}(X)) = \frac{5\kappa}{6} \sum_{j=0}^{\infty} \binom{\lfloor \frac{r}{2} \rfloor + j}{j} \int_0^X \frac{\theta_{n(j)}(H)}{H^{1/6}} \cdot \mathbb{E}_{\lfloor \frac{r}{2} \rfloor, j}^{n(j)}(H) dH + \left( \sum_{j=0}^{\infty} \theta_{r+2j}(X) \right) \cdot O \left( X^{1/2} \right).$$

where the error term satisfies  $\left(\sum_{j=0}^{\infty} \theta_{r+2j}(X)\right) \cdot O(X^{1/2}) = O(X^{1/2})$ , and  $\mathbb{E}_{\lfloor \frac{r}{2} \rfloor, j}^{n(j)}(H)$  is the expected value defined in Remark 5.12.

**Remark 6.3.** If we assume  $H_A$ ,  $H_B$ , and Conjectures 4.5 and 5.19, and in addition we assume that the random variables  $Y_1, \dots, Y_{\lfloor n(j)/2 \rfloor}$  are independent, then we would have

$$\begin{aligned} \pi_{\mathcal{R}_r}(X) &\approx \frac{5\kappa}{6} \sum_{j=0}^{\infty} \binom{\lfloor \frac{r}{2} \rfloor + j}{j} \int_0^X \frac{\theta_{n(j)}(H)}{H^{1/6}} \cdot \rho_{n(j)}(H)^{\lfloor r/2 \rfloor} (1 - \rho_{n(j)}(H))^j dH. \\ &\approx \frac{5\kappa}{6} \sum_{j=0}^{\infty} \binom{\lfloor \frac{r}{2} \rfloor + j}{j} \int_0^X \frac{s_{n(j)} \cdot (D_{n(j)})^{\lfloor \frac{r}{2} \rfloor} \cdot (H^{f_{n(j)}} - D_{n(j)})^j}{(1 + C_{n(j)} H^{-e_{n(j)}}) \cdot H^{1/6 + (\lfloor \frac{r}{2} \rfloor + j) \cdot f_{n(j)}}} dH. \end{aligned}$$

If we simplify this expression further by just retaining the highest order term (and for now assume  $r \geq 2$ ). We obtain:

$$\begin{aligned} \pi_{\mathcal{R}_r}(X) &\approx \frac{5\kappa}{6} \sum_{j=0}^{\infty} \binom{\lfloor \frac{r}{2} \rfloor + j}{j} \cdot s_{n(j)} \cdot (D_{n(j)})^{\lfloor \frac{r}{2} \rfloor} \int_0^X \frac{1}{H^{1/6 + \lfloor \frac{r}{2} \rfloor \cdot f_{n(j)}}} dH \\ &\approx \frac{5\kappa}{6} \sum_{j=0}^{\infty} \binom{\lfloor \frac{r}{2} \rfloor + j}{j} \cdot s_{n(j)} \cdot (D_{n(j)})^{\lfloor \frac{r}{2} \rfloor} \cdot \frac{X^{5/6 - \lfloor \frac{r}{2} \rfloor \cdot f_{n(j)}}}{5/6 - \lfloor \frac{r}{2} \rfloor \cdot f_{n(j)}}. \end{aligned}$$

In particular, if there is  $j \geq 0$  such that  $\lfloor \frac{r}{2} \rfloor \cdot f_{n(j)} < 5/6$ , then there are infinitely many elliptic curves with rank  $r$  (and Selmer rank  $n(j)$ ).

In our next result, we use Theorem 6.1 to write formulas for the contribution in rank  $r = 1, \dots, 5$  coming from Selmer ranks  $n = 1, \dots, 5$ .

**Corollary 6.4.** *If we assume  $H_A$ ,  $H_B$ , and Conjectures 4.5 and 5.19, then the formulas in Corollary 5.15 imply approximations of  $\pi_{\mathcal{R}_r \cap \mathcal{S}_n}(X)$  as given in Table 13, for  $1 \leq r \leq n \leq 5$  and  $r \equiv n \pmod{2}$ .*

**Remark 6.5.** Using the formulas given by Corollary 6.4 and Table 14, we can give approximations of  $\pi_{\mathcal{R}_r}(X)$ . For instance,

$$\begin{aligned} \pi_{\mathcal{R}_1}(X) &\approx \pi_{\mathcal{R}_1 \cap \mathcal{S}_1}(X) + \pi_{\mathcal{R}_1 \cap \mathcal{S}_3}(X) + \pi_{\mathcal{R}_1 \cap \mathcal{S}_5}(X), \\ \pi_{\mathcal{R}_2}(X) &\approx \pi_{\mathcal{R}_2 \cap \mathcal{S}_2}(X) + \pi_{\mathcal{R}_2 \cap \mathcal{S}_4}(X), \quad \pi_{\mathcal{R}_3}(X) \approx \pi_{\mathcal{R}_3 \cap \mathcal{S}_3}(X) + \pi_{\mathcal{R}_3 \cap \mathcal{S}_5}(X), \\ \pi_{\mathcal{R}_4}(X) &\approx \pi_{\mathcal{R}_4 \cap \mathcal{S}_4}(X), \quad \pi_{\mathcal{R}_5}(X) \approx \pi_{\mathcal{R}_5 \cap \mathcal{S}_5}(X). \end{aligned}$$

We have used SageMath to numerically integrate and compute said approximations, and we have graphed the results in Figures 16 (for  $r = 1, 2, 3$ ) and 17 (for  $r = 4, 5$ ). In Table 14 we have included the values of  $\pi_{\mathcal{R}_r}(2.7 \cdot 10^{10})$  according to the data, the values of our approximation, the error, and the relative error (as a percentage of the actual value), and also  $s_r \cdot (2.7 \cdot 10^{10})^{1/2}$ , which is, approximately, the size of the error as expected from Corollary 6.2.

Next, using the formulas from Corollary 6.4, we can estimate the rate of growth of our rank counting functions  $\pi_{\mathcal{R}_r \cap \mathcal{S}_n}(X)$ . For instance:

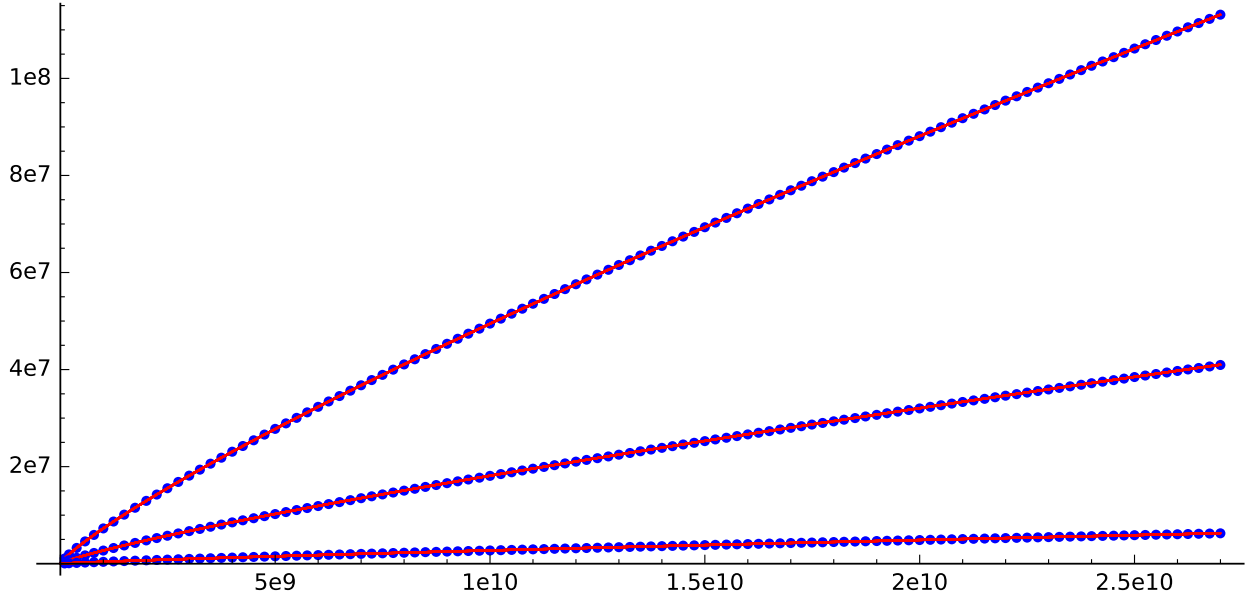


FIGURE 16. Values of  $\pi_{\mathcal{R}_r}(X)$  from the BHKSSW database (blue dots) for  $r = 1, 2, 3$ , and the approximations given in Remark 6.5 (in red).

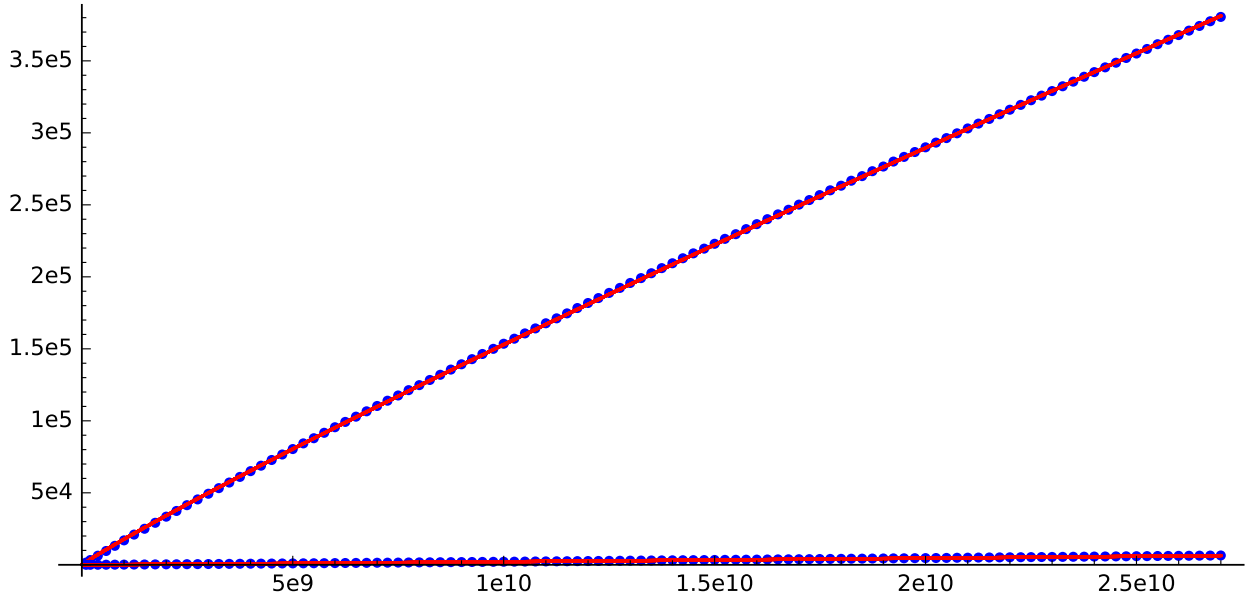


FIGURE 17. Values of  $\pi_{\mathcal{R}_r}(X)$  from the BHKSSW database (blue dots) for  $r = 4, 5$ , and the approximations given in Remark 6.5 (in red).

$\pi_{\mathcal{R}_1 \cap \mathcal{S}_1}(X)$	$\frac{5\kappa}{6} \int_0^X \frac{\theta_1(H)}{H^{1/6}} dH \approx \frac{5\kappa}{6} \int_0^X \frac{s_1}{(1 + C_1 H^{-e_1}) H^{1/6}} dH$
$\pi_{\mathcal{R}_1 \cap \mathcal{S}_3}(X)$	$\frac{5\kappa}{6} \int_0^X \frac{\theta_3(H)}{H^{1/6}} \cdot (1 - \rho_3(H)) dH \approx \frac{5\kappa}{6} \int_0^X \frac{s_3 \cdot (H^{f_3} - D_3)}{(1 + C_3 H^{-e_3}) H^{1/6+f_3}} dH.$
$\pi_{\mathcal{R}_1 \cap \mathcal{S}_5}(X)$	$\frac{5\kappa}{6} \int_0^X \frac{\theta_5(H)}{H^{1/6}} \cdot (1 - \rho_5(H))^2 dH \approx \frac{5\kappa}{6} \int_0^X \frac{s_5 \cdot (H^{f_5} - D_5)^2}{(1 + C_5 H^{-e_5}) H^{1/6+2f_5}} dH.$
$\pi_{\mathcal{R}_2 \cap \mathcal{S}_2}(X)$	$\frac{5\kappa}{6} \int_0^X \frac{\theta_2(H)}{H^{1/6}} \cdot \rho_2(H) dH \approx \frac{5\kappa}{6} \int_0^X \frac{s_2 \cdot D_2}{(1 + C_2 H^{-e_2}) H^{1/6+f_2}} dH.$
$\pi_{\mathcal{R}_2 \cap \mathcal{S}_4}(X)$	$\begin{aligned} & \frac{10\kappa}{6} \int_0^X \frac{\theta_4(H)}{H^{1/6}} \cdot (\rho_4(H)(1 - \rho_4(H)) - C_{1,1}^4(X)) dH \\ & \approx \frac{10\kappa}{6} \int_0^X \frac{s_4 \cdot (-(D_4)^2 + D_4 \cdot H^{f_4} + 0.025 \cdot (H^{f_4})^2)}{(1 + C_4 H^{-e_4}) H^{1/6+2f_4}} dH. \end{aligned}$
$\pi_{\mathcal{R}_3 \cap \mathcal{S}_3}(X)$	$\frac{5\kappa}{6} \int_0^X \frac{\theta_3(H)}{H^{1/6}} \cdot \rho_3(H) dH \approx \frac{5\kappa}{6} \int_0^X \frac{s_3 \cdot D_3}{(1 + C_3 H^{-e_3}) H^{1/6+f_3}} dH.$
$\pi_{\mathcal{R}_3 \cap \mathcal{S}_5}(X)$	$\begin{aligned} & \frac{10\kappa}{6} \int_0^X \frac{\theta_5(H)}{H^{1/6}} \cdot (\rho_5(H)(1 - \rho_5(H)) - C_{1,1}^5(X)) dH \\ & \approx \frac{10\kappa}{6} \int_0^X \frac{s_5 \cdot D_5 \cdot (H^{f_5} - D_5)}{(1 + C_5 H^{-e_5}) H^{1/6+2f_5}} dH. \end{aligned}$
$\pi_{\mathcal{R}_4 \cap \mathcal{S}_4}(X)$	$\frac{5\kappa}{6} \int_0^X \frac{\theta_4(H)}{H^{1/6}} \cdot (\rho_4(H)^2 + C_{1,1}^4(X)) dH \approx \frac{5\kappa}{6} \int_0^X \frac{s_4 \cdot ((D_4)^2 - 0.025 \cdot (H^{f_4})^2)}{(1 + C_4 H^{-e_4}) H^{1/6+2f_4}} dH.$
$\pi_{\mathcal{R}_5 \cap \mathcal{S}_5}(X)$	$\frac{5\kappa}{6} \int_0^X \frac{\theta_5(H)}{H^{1/6}} \cdot (\rho_5(H)^2 + C_{1,1}^5(X)) dH \approx \frac{5\kappa}{6} \int_0^X \frac{s_5 \cdot (D_5)^2}{(1 + C_5 H^{-e_5}) H^{1/6+2f_5}} dH.$

TABLE 13. Approximate values of  $\pi_{\mathcal{R}_r \cap \mathcal{S}_n}(X)$  for  $1 \leq r \leq n \leq 5$  and  $r \equiv n \pmod 2$ .

**Corollary 6.6.** *If we assume  $H_A$ ,  $H_B$ , and Conjectures 4.5 and 5.19, then there are explicit computable positive constants  $\lambda_r$  and  $h_r$ , for  $n = 1, 2, 3$  such that*

$$\mathbb{E}(\pi_{\mathcal{R}_1 \cap \mathcal{S}_1}(X)) = \lambda_1 + \kappa s_1 X^{5/6} \cdot \sum_{m=0}^{\infty} \frac{(-C_1)^m}{1 - (6/5) \cdot m e_1} X^{-m e_1} + O(X^{1/2}),$$

	$r = 1$	2	3	4	5
$\pi_{\mathcal{R}_r}(2.7 \cdot 10^{10})$	113128929	40949289	6259157	380519	6481
Approximate value	113133971	41005107	6273138	381272	6438
Error	5042	55818	13981	753	43
Error %	0.004456	0.136310	0.223368	0.197887	0.663477
Predicted error $\approx s_r \cdot X^{1/2}$	68848.72	45942.96	13112.47	1749.97	111.73

TABLE 14. Values of  $\pi_{\mathcal{R}_r}(2.7 \cdot 10^{10})$  from the BHKSWW database, the approximate values (rounded to the closest integer) given by numerical integration of the formulas in Table 13 and Remark 6.5, the absolute error, the error as a percentage of the actual value of  $\pi_{\mathcal{R}_r}$ , and the size of the predicted error  $s_r \cdot (2.7 \cdot 10^{10})^{1/2}$  from Corollary 6.2.

$$\mathbb{E}(\pi_{\mathcal{R}_2 \cap \mathcal{S}_2}(X)) = \lambda_2 + \kappa s_2 D_2 X^{5/6-f_2} \cdot \sum_{m=0}^{\infty} \frac{(-C_2)^m}{1 - (6/5) \cdot (f_2 + m e_2)} X^{-m e_2} + O(X^{1/2}),$$

$$\mathbb{E}(\pi_{\mathcal{R}_3 \cap \mathcal{S}_3}(X)) = \lambda_3 + \kappa s_3 D_3 X^{5/6-f_3} \cdot \sum_{m=0}^{\infty} \frac{(-C_3)^m}{1 - (6/5) \cdot (f_3 + m e_3)} X^{-m e_3} + O(X^{1/2}),$$

for any  $X \geq h_r$ .

*Proof.* For each  $r = 1, 2, 3$ , let  $h_r > 0$  be the smallest natural number such that  $|C_r h_r^{-e_r}| < 1$ . Then,

$$\pi_{\mathcal{R}_r \cap \mathcal{S}_r}(X) = \pi_{\mathcal{R}_r \cap \mathcal{S}_r}(h_r) + \pi_{\mathcal{R}_r \cap \mathcal{S}_r}([h_r, X])$$

and, by Corollary 6.2 we have

$$\mathbb{E}(\pi_{\mathcal{R}_r \cap \mathcal{S}_r}([h_0, X])) = \frac{5\kappa}{6} \binom{\lfloor \frac{r}{2} \rfloor}{0} \int_{h_0}^X \frac{s_r \cdot \mathbb{E}_{\lfloor \frac{r}{2} \rfloor, 0}^r(H)}{(1 + C_r H^{-e_r}) \cdot H^{1/6}} dH + O(X^{1/2}).$$

Further, since  $|C_r h_r^{-e_r}| < 1$ , we can write

$$\frac{1}{1 + C_r H^{-e_r}} = \sum_{m=0}^{\infty} (-C_r)^m H^{-m e_r}$$

for any  $H \geq h_r$ . Now,  $\mathbb{E}_{\lfloor \frac{r}{2} \rfloor, 0}^r(H) = 1$  for  $r = 1$ , and by Corollary 5.15, we have  $\mathbb{E}_{\lfloor \frac{r}{2} \rfloor, 0}^r(H) = \rho_r(X)$  for  $r = 2, 3$ . Further, assuming  $H_B$  we have  $\rho_n(X) = D_n/X^{f_n}$ . Putting everything together we

obtain, for instance, the following approximation formula for  $\mathbb{E}(\pi_{\mathcal{R}_1 \cap \mathcal{S}_1}(X))$

$$\begin{aligned}
&= \pi_{\mathcal{R}_1 \cap \mathcal{S}_1}(h_1) + \pi_{\mathcal{R}_1 \cap \mathcal{S}_1}([h_1, X]) \\
&= \pi_{\mathcal{R}_1 \cap \mathcal{S}_1}(h_1) + \frac{5\kappa}{6} \int_{h_1}^X \frac{\theta_1(H)}{H^{1/6}} dH + O(X^{1/2}) \\
&= \pi_{\mathcal{R}_1 \cap \mathcal{S}_1}(h_1) + \frac{5\kappa s_1}{6} \int_{h_1}^X \sum_{m=0}^{\infty} (-C_r)^m H^{-\frac{1}{6} - me_1} dH + O(X^{1/2}) \\
&= \pi_{\mathcal{R}_1 \cap \mathcal{S}_1}(h_1) - \left( \kappa s_1 h_1^{5/6} \cdot \sum_{m=0}^{\infty} \frac{(-C_1)^m}{1 - (6/5) \cdot me_1} h_1^{-me_1} \right) + \kappa s_1 X^{5/6} \cdot \sum_{m=0}^{\infty} \frac{(-C_1)^m}{1 - (6/5) \cdot me_1} X^{-me_1} \\
&= \lambda_1 + \kappa s_1 X^{5/6} \cdot \sum_{m=0}^{\infty} \frac{(-C_1)^m}{1 - (6/5) \cdot me_1} X^{-me_1} + O(X^{1/2}),
\end{aligned}$$

with  $\lambda_1 = \pi_{\mathcal{R}_1 \cap \mathcal{S}_1}(h_0) - \left( \kappa s_1 h_1^{5/6} \cdot \sum_{m=0}^{\infty} \frac{(-C_1)^m}{1 - (6/5) \cdot me_1} h_1^{-me_1} \right)$ , and we derive formulas for  $r = 2$  and  $r = 3$  in a similar manner.  $\square$

## 7. PREDICTING THE AVERAGE RANK

In this section we shall assume hypotheses  $H_A$  and  $H_B$  and estimate the average rank of all elliptic curves of height  $\leq X$ :

$$\text{AvgRank}_{\mathcal{E}}(X) = \frac{\sum_{E \in \mathcal{E}(X)} \text{rank}(E(\mathbb{Q}))}{\pi_{\mathcal{E}}(X)}$$

We quote here the average rank conjecture as in [22] (see [11] for Goldfeld's version for quadratic twists).

**Conjecture 7.1.** *Fix a global field  $k$ . Asymtotically, 50% of elliptic curves over  $k$  have rank 0, and 50% have rank 1. Moreover, the average rank is  $1/2$ .*

Here we consider the average rank contributions from the subsets of elliptic curves of each Selmer rank  $n \geq 1$ :

$$\text{AvgRank}_{\mathcal{S}_n}(X) = \frac{\sum_{E \in \mathcal{S}_n(X)} \text{rank}(E(\mathbb{Q}))}{\pi_{\mathcal{E}}(X)},$$

and later we will put them together to estimate the total average rank.

**Theorem 7.2.** *Assume  $H_A$  and  $H_B$ , and let  $n \geq 1$  be fixed. Then, the expected value of  $\text{AvgRank}_{\mathcal{S}_n}(X)$  is given by*

$$\frac{5\kappa}{6\pi_{\mathcal{E}}(X)} \cdot \int_0^X \frac{\theta_n(H)}{H^{1/6}} \left( (n \bmod 2) + 2 \left\lfloor \frac{n}{2} \right\rfloor \rho_n(H) \right) dH + \theta_n(X) \cdot O(X^{-1/3}).$$

Moreover, the error in approximating  $\text{AvgRank}_{\mathcal{S}_n}(X)$  by its expected value is approximately given by

$$\sqrt{\frac{5\kappa \lfloor n/2 \rfloor}{6\pi_{\mathcal{E}}(X)^2} \int_0^X \frac{\theta_n(H)}{H^{1/6}} (\rho_n(H)(1 - \rho_n(H)) + (\lfloor n/2 \rfloor - 1)C_{1,1}^n(H)) dH + O(X^{-7/6})}.$$



*Proof.* We compute the expected value of the average rank as follows:

$$\begin{aligned}\mathbb{E}(\text{AvgRank}_{\mathcal{S}_n}(X)) &= \mathbb{E}\left(\frac{\sum_{E \in \mathcal{S}_n(X)} \text{rank}(E(\mathbb{Q}))}{\pi_{\mathcal{E}}(X)}\right) = \frac{1}{\pi_{\mathcal{E}}(X)} \mathbb{E}\left(\sum_{E \in \mathcal{S}_n(X)} \text{rank}(E(\mathbb{Q}))\right) \\ &= \frac{1}{\pi_{\mathcal{E}}(X)} \cdot \left(\sum_{E \in \mathcal{S}_n(X)} (n \bmod 2) + 2 \left\lfloor \frac{n}{2} \right\rfloor \rho_n(\text{ht}(E))\right)\end{aligned}$$

by Corollary 5.10. In particular, Theorem 3.1 and  $H_A$  imply

$$\begin{aligned}&= \frac{1}{\pi_{\mathcal{E}}(X)} \cdot \sum_{H=1}^X \sum_{E \in \mathcal{S}_n([H, H])} (n \bmod 2) + 2 \left\lfloor \frac{n}{2} \right\rfloor \rho_n(H) \\ &= \frac{1}{\pi_{\mathcal{E}}(X)} \cdot \sum_{H=1}^X \pi_{\mathcal{S}_n}([H, H]) \cdot \left((n \bmod 2) + 2 \left\lfloor \frac{n}{2} \right\rfloor \rho_n(H)\right) \\ &= \frac{1}{\pi_{\mathcal{E}}(X)} \cdot \left(\frac{5\kappa}{6} \int_0^X \frac{\theta_n(H)}{H^{1/6}} \left((n \bmod 2) + 2 \left\lfloor \frac{n}{2} \right\rfloor \rho_n(H)\right) dH + \theta_n(X) \cdot O(X^{1/2})\right) \\ &= \frac{5\kappa}{6\pi_{\mathcal{E}}(X)} \cdot \int_0^X \frac{\theta_n(H)}{H^{1/6}} \left((n \bmod 2) + 2 \left\lfloor \frac{n}{2} \right\rfloor \rho_n(H)\right) dH + \theta_n(X) \cdot O(X^{-1/3}),\end{aligned}$$

where we have used Proposition 4.9 and Corollary 3.4 for the estimate  $\pi_{\mathcal{E}}(X) = O(X^{5/6})$ . Moreover, by Corollary 5.10, the standard error in the approximation of the average by the expected value is given by

$$\begin{aligned}&\frac{1}{\pi_{\mathcal{E}}(X)} \sqrt{\left\lfloor \frac{n}{2} \right\rfloor \sum_{E \in \mathcal{S}_n(X)} \rho_n(\text{ht}(E))(1 - \rho_n(\text{ht}(E))) + (\lfloor n/2 \rfloor - 1) C_{1,1}^n(\text{ht}(E))}\right. \\ &= \frac{1}{\pi_{\mathcal{E}}(X)} \sqrt{\frac{5\kappa \lfloor n/2 \rfloor}{6} \int_0^X \frac{\theta_n(H)}{H^{1/6}} (\rho_n(H)(1 - \rho_n(H)) + (\lfloor n/2 \rfloor - 1) C_{1,1}^n(H)) dH + O(X^{1/2})}, \\ &= \sqrt{\frac{5\kappa \lfloor n/2 \rfloor}{6\pi_{\mathcal{E}}(X)^2} \int_0^X \frac{\theta_n(H)}{H^{1/6}} (\rho_n(H)(1 - \rho_n(H)) + (\lfloor n/2 \rfloor - 1) C_{1,1}^n(H)) dH + O(X^{-7/6})}.\end{aligned}$$

□

**Remark 7.3.** Let  $h_n$  be the smallest positive integer such that  $|C_n h_n^{-e_n}| < 1$ . If we assume Conjectures 4.5 and 5.19, then  $\text{AvgRank}_{\mathcal{S}_n}(X)$  is approximately given by

$$\begin{aligned}&\approx \frac{5\kappa}{6\pi_{\mathcal{E}}(X)} \cdot \int_0^X \frac{\theta_n(H)}{H^{1/6}} \left((n \bmod 2) + 2 \left\lfloor \frac{n}{2} \right\rfloor \rho_n(H)\right) dH \\ &\approx \frac{(5/6)\kappa s_n}{\pi_{\mathcal{E}}(X)} \cdot \int_0^X \frac{1}{H^{1/6}(1 + C_n H^{-e_n})} \left((n \bmod 2) + 2 \left\lfloor \frac{n}{2} \right\rfloor \frac{D_n}{H^{f_n}}\right) dH \\ &\approx \frac{(5/6)\kappa s_n}{\pi_{\mathcal{E}}(X)} \cdot \left(\mu_n + \int_{h_n}^X \sum_{m=0}^{\infty} (-C_n)^m H^{-1/6-m e_n} \left((n \bmod 2) + 2 \left\lfloor \frac{n}{2} \right\rfloor \frac{D_n}{H^{f_n}}\right) dH\right)\end{aligned}$$

where  $\mu_n = \int_0^{h_n} \sum_{m=0}^{\infty} (-C_n)^m H^{-1/6-me_n} \left( (n \bmod 2) + 2 \left\lfloor \frac{n}{2} \right\rfloor \frac{D_n}{H f_n} \right) dH$ . Thus,

$$\begin{aligned} &\approx \frac{(5/6)\kappa s_n}{\pi_{\mathcal{E}}(X)} \cdot \left( \mu_n - \bar{n} \sum_{m=0}^{\infty} \frac{(-C_n)^m}{5/6 - me_n} (h_n)^{5/6-me_n} - 2 \left\lfloor \frac{n}{2} \right\rfloor \sum_{m=0}^{\infty} \frac{D_n(-C_n)^m}{5/6 - f_n - me_n} (h_n)^{5/6-f_n-me_n} \right) \\ &+ \frac{(5/6)\kappa s_n}{\pi_{\mathcal{E}}(X)} \cdot \left( \bar{n} \sum_{m=0}^{\infty} \frac{(-C_n)^m}{5/6 - me_n} X^{5/6-me_n} + 2 \left\lfloor \frac{n}{2} \right\rfloor \sum_{m=0}^{\infty} \frac{D_n(-C_n)^m}{5/6 - f_n - me_n} X^{5/6-f_n-me_n} \right), \end{aligned}$$

where we have abbreviated  $\bar{n} = (n \bmod 2)$ , and below we shall write  $\tau_n$  for the contents inside the first parenthesis, i.e.,  $\tau_n = \mu_n - \bar{n} \sum_{m=0}^{\infty} \dots - 2 \left\lfloor \frac{n}{2} \right\rfloor \sum_{m=0}^{\infty} \dots$

$$\begin{aligned} &\approx \frac{\kappa s_n X^{5/6}}{\pi_{\mathcal{E}}(X)} \cdot \left( \frac{\tau_n}{X^{5/6}} + \sum_{m=0}^{\infty} \left( \frac{(n \bmod 2)(-C_n)^m}{1 - (6/5)me_n} + X^{-f_n} \frac{2 \left\lfloor \frac{n}{2} \right\rfloor D_n(-C_n)^m}{1 - (6/5)(f_n + me_n)} \right) X^{-me_n} \right) \\ &\approx s_n \cdot \left( \frac{\tau_n}{X^{5/6}} + \sum_{m=0}^{\infty} \left( \frac{(n \bmod 2)(-C_n)^m}{1 - (6/5)me_n} + X^{-f_n} \frac{2 \left\lfloor \frac{n}{2} \right\rfloor D_n(-C_n)^m}{1 - (6/5)(f_n + me_n)} \right) X^{-me_n} \right). \end{aligned}$$

Hence, we obtain the following result about the average rank of elliptic curves.

**Corollary 7.4.** *If we assume  $H_A$  and  $H_B$ , and Conjectures 4.5 and 5.19, then there are constants  $\tau_n$  such that the expected value of  $\text{AvgRank}_{\mathcal{E}}(X)$  is given by*

$$\begin{aligned} &= \sum_{n=1}^{\infty} \text{AvgRank}_{\mathcal{S}_n}(X) \\ &\approx \sum_{n=1}^{\infty} s_n \cdot \left( \frac{\tau_n}{X^{5/6}} + \sum_{m=0}^{\infty} \left( \frac{(n \bmod 2)(-C_n)^m}{1 - (6/5)me_n} + X^{-f_n} \frac{2 \left\lfloor \frac{n}{2} \right\rfloor D_n(-C_n)^m}{1 - (6/5)(f_n + me_n)} \right) X^{-me_n} \right). \end{aligned}$$

with standard error  $\leq \frac{\sum_{n=2}^{\infty} \sqrt{\lfloor n/2 \rfloor \cdot (\lfloor n/2 \rfloor - 3/4) \cdot s_n}}{\sqrt{\kappa} X^{5/12}}$ . In particular,

$$\lim_{X \rightarrow \infty} \text{AvgRank}_{\mathcal{E}}(X) = \sum_{k=0}^{\infty} s_{2k+1} = \frac{1}{2},$$

with standard error going to 0 as  $X \rightarrow \infty$ .

*Proof.* The approximation of the average rank is an immediate consequence of our approximation of the contribution to the average rank coming from each Selmer rank  $n$  given in Remark 7.3. From the approximation, it follows that

$$\lim_{X \rightarrow \infty} \text{AvgRank}_{\mathcal{E}}(X) \approx \sum_{n=1}^{\infty} s_n \cdot (n \bmod 2) = \sum_{k=0}^{\infty} s_{2k+1}.$$

Finally, we point out that, by Proposition 2.6 of [22], the values  $s_n$  have a generating function

$$\sum_{n \geq 0} s_n z^n = \prod_{i=0}^{\infty} \frac{1 + 2^{-i} z}{1 + 2^{-i}}.$$

In particular, for  $z = 1$  we obtain that  $\sum_n s_n = 1$ , for  $z = -1$  we obtain that  $\sum_n (-1)^n s_n = 0$ , and therefore  $\sum_k s_{2k+1} = \sum_{n \equiv 1 \pmod 2} s_n = \frac{1}{2} (\sum_n s_n - \sum_n (-1)^n s_n) = \frac{1}{2}$ . Let us now estimate the error in the approximation of  $\text{AvgRank}_{\mathcal{S}_n}(X)$  using Corollary 5.10:

$$\begin{aligned} & \frac{1}{\pi_{\mathcal{E}}(X)} \sqrt{\lfloor n/2 \rfloor \sum_{E \in \mathcal{S}_n(X)} \rho_n(\text{ht}(E))(1 - \rho_n(\text{ht}(E))) + (\lfloor n/2 \rfloor - 1)C_{1,1}^n(\text{ht}(E))} \\ & \approx \frac{1}{\pi_{\mathcal{E}}(X)} \sqrt{\frac{5\kappa \lfloor n/2 \rfloor}{6} \int_1^X \frac{\theta_n(H)}{H^{1/6}} (\rho_n(H)(1 - \rho_n(H)) + (\lfloor n/2 \rfloor - 1)C_{1,1}^n(H)) dH} \\ & \approx \frac{1}{\pi_{\mathcal{E}}(X)} \sqrt{\frac{5\kappa \lfloor n/2 \rfloor s_n}{6} \int_1^X \frac{1}{H^{1/6}(1 + C_n H^{-e_n})} \left( \frac{D_n}{H f_n} \left( 1 - \frac{D_n}{H f_n} \right) + (\lfloor n/2 \rfloor - 1)C_{1,1}^n(H) \right) dH}. \end{aligned}$$

Recall that the covariance coefficient  $C_{1,1}^n(X)$  is given by  $\mathbb{E}(Y_1 Y_2) - \mathbb{E}(Y_1)\mathbb{E}(Y_2)$ , and since the random variables  $Y_i$  take only the values 0, 1, we have  $0 \leq \mathbb{E}(Y_1) = \rho_n(X) \leq 1$ . In particular,  $|C_{1,1}^n(X)| \leq 1$ . Also, notice that  $y(1 - y)$  in the interval  $[0, 1]$  obtains the maximum value of  $1/4$  at  $y = 1/2$ . Thus,

$$\begin{aligned} & \approx \frac{1}{\pi_{\mathcal{E}}(X)} \sqrt{\frac{5\kappa \lfloor n/2 \rfloor s_n}{6} \int_1^X \frac{1}{H^{1/6}(1 + C_n H^{-e_n})} \left( \frac{D_n}{H f_n} \left( 1 - \frac{D_n}{H f_n} \right) + (\lfloor n/2 \rfloor - 1)C_{1,1}^n(H) \right) dH} \\ & \leq \frac{1}{\pi_{\mathcal{E}}(X)} \sqrt{\frac{5\kappa \lfloor n/2 \rfloor s_n}{6} \int_1^X \frac{1/4 + (\lfloor n/2 \rfloor - 1)}{H^{1/6}} dH} \\ & \leq \frac{1}{\pi_{\mathcal{E}}(X)} \sqrt{\kappa \lfloor n/2 \rfloor s_n (\lfloor n/2 \rfloor - 3/4) X^{5/6}} \\ & = \frac{\sqrt{\lfloor n/2 \rfloor (\lfloor n/2 \rfloor - 3/4) s_n}}{\sqrt{\kappa} X^{5/12}}. \end{aligned}$$

Thus, the standard error in the approximation of  $\text{AvgRank}_{\mathcal{E}}(X)$  is bounded by

$$\frac{\sum_{n=2}^{\infty} \sqrt{\lfloor n/2 \rfloor (\lfloor n/2 \rfloor - 3/4) s_n}}{\sqrt{\kappa} X^{5/12}}$$

It remains to show that  $\sum_{n=2}^{\infty} \sqrt{\lfloor n/2 \rfloor (\lfloor n/2 \rfloor - 3/4) s_n}$  is convergent. Let us define  $t_1 = s_1$  and

$$t_n = \frac{t_1}{2^{\frac{n(n-1)}{2} - 1}}$$

for  $n \geq 2$ . Then, the definition of  $s_n$  implies that  $s_n \leq t_n$ , and therefore,

$$\sum_{n=2}^N \sqrt{\lfloor n/2 \rfloor (\lfloor n/2 \rfloor - 3/4) s_n} \leq \sum_{n=2}^N \frac{n}{2} \sqrt{s_n} \leq \sum_{n=2}^N \frac{n}{2} \sqrt{t_n} \leq \sum_{n=2}^N \frac{n}{2} \frac{\sqrt{t_1}}{2^{\frac{n(n-1)-2}{4}}} \leq \sum_{n=2}^N \frac{\sqrt{s_1} \cdot n}{2^{\frac{n(n-1)+2}{4}}}$$

for any  $N$ , and therefore  $\sum_{n=2}^{\infty} \sqrt{\lfloor n/2 \rfloor (\lfloor n/2 \rfloor - 3/4) s_n}$  is convergent. Thus, the error goes to 0 as  $X \rightarrow \infty$ , as desired.  $\square$

**Remark 7.5.** Using SageMath, in Figure 18 we have plotted values of  $\text{AvgRank}_{\mathcal{E}}(X)$  from the BHKSSW database, and (via numerical integration) the sum of the approximations given in Theorem

7.2 of  $\text{AvgRank}_{\mathcal{S}_n}(X)$  for  $n = 1, \dots, 5$ . According to the database, we have

$$\text{AvgRank}_{\mathcal{E}}(2.7 \cdot 10^{10}) = 0.90197580$$

while our approximation gives 0.90244770. Thus, the absolute error is 0.00047189, which represents a 0.0523% of the true value.

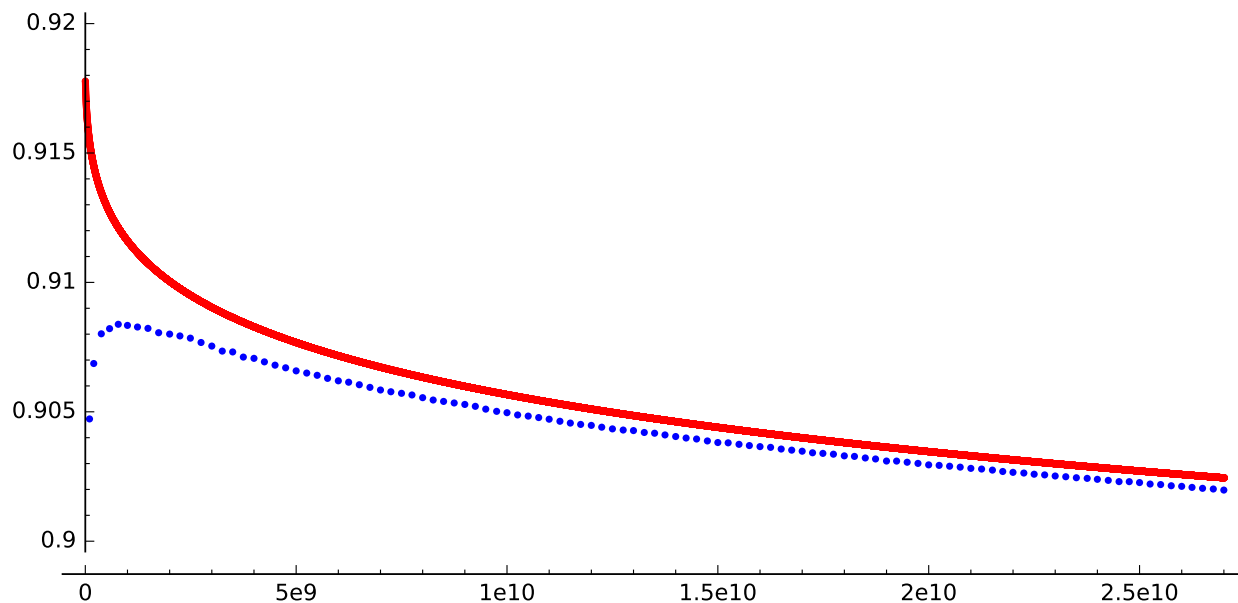


FIGURE 18. Values of  $\text{AvgRank}_{\mathcal{E}}(X)$  from the BHKSSW database (blue dots), and the approximation given in Corollary 7.4 (in red).

## REFERENCES

- [1] J. S. Balakrishnan, W. Ho, N. Kaplan, S. Spicer, W. Stein, J. Weigandt, *Databases of elliptic curves ordered by height and distributions of Selmer groups and ranks*, to appear in ANTS XII. [1](#), [1.6](#), [3](#), [4.7](#), [5.23](#)
- [2] B. Bektemirov, B. Mazur, W. Stein, M. Watkins, *Average ranks of elliptic curves: tension between data and conjecture*, Bull. (New Series) Amer. Math. Soc., Vol. 44, No. 2, April 2007, 233-254. [6](#)
- [3] W. Bosma, J. Cannon, and C. Playoust, *The Magma algebra system. I. The user language*, J. Symbolic Comput., **24** (1997), 235-265.
- [4] A. Brumer, *The average rank of elliptic curves I*, Invent. Math. 109 (1992), no. 3, 445-472. [3](#), [3](#)
- [5] J. W. S. Cassels, *Arithmetic on curves of genus 1. IV. Proof of the Hauptvermutung*, Journal für die reine und angewandte Mathematik, 211: 95-112. [5](#)
- [6] J. E. Cremona, *Elliptic curve data*, database available at <http://homepages.warwick.ac.uk/~masgaj/ftp/data/>.
- [7] J. E. Cremona, B. Mazur, *Visualizing Elements in the ShafarevichTate Group*, Experiment. Math. 9:1 (2000), 13-28. [5](#)
- [8] C. Delaunay, *Heuristics on Tate-Shafarevich groups of elliptic curves defined over  $\mathbb{Q}$* , Experiment. Math. 10 (2001), no. 2, 191-196. [5](#)
- [9] C. Delaunay, *Moments of the orders of Tate-Shafarevich groups*, Int. J. Number Theory 1 (2005), no. 2, 243-264. [5](#)

- [10] A. Dujella, Website: <https://web.math.pmf.unizg.hr/~duje/tors/tors.html>, *High rank elliptic curves with prescribed torsion*. [1](#)
- [11] D. Goldfeld, *Conjectures on elliptic curves over quadratic fields*, Number theory, Carbondale 1979 (Proc. Southern Illinois Conf., Southern Illinois Univ., Carbondale, Ill., 1979), Lecture Notes in Math., vol. 751, Springer, Berlin, 1979, pp. 108-118. [7](#)
- [12] G.H. Hardy, E. M. Wright, *An Introduction to the Theory of Numbers*, 5th ed. Oxford, England: Clarendon Press, 1979.
- [13] R. Harron, A. Snowden, *Counting elliptic curves with prescribed torsion*, to appear in Journal für die reine und angewandte Mathematik. [1](#)
- [14] D. R. Heath-Brown, *The size of Selmer groups for the congruent number problem*, Invent. Math. 111 (1993), no. 1, 171-195. [4](#)
- [15] D. R. Heath-Brown, *The size of Selmer groups for the congruent number problem. II*, Invent. Math. 118 (1994), no. 2, 331-370. [4](#)
- [16] D. M. Kane, *On the ranks of the 2-Selmer groups of twists of a given elliptic curve*, Algebra Number Theory 7 (2013), 1253-1279. [4](#)
- [17] M. A. Kenku, *On the number of  $\mathbb{Q}$ -isomorphism classes of elliptic curves in each  $\mathbb{Q}$ -isogeny class*, J. Number Theory **15** (1982), 199-202. [4](#) [2](#)
- [18] D. S. Kubert, *Universal bounds on the torsion of elliptic curves*, Compositio Math. 38 (1979), no. 1, 121-128. [1](#)
- [19] B. Mazur, *Modular curves and the Eisenstein ideal*, Inst. Hautes Études Sci. Publ. Math. **47** (1977), 33-186. [1](#)
- [20] B. Mazur, *Rational isogenies of prime degree*, Invent. Math. **44** (1978), 129-162. [1](#)
- [21] J. Park, B. Poonen, J. Voight, M. M. Wood, *A heuristic for boundedness of ranks of elliptic curves*, submitted. [1](#), [3](#), [5](#), [6](#)
- [22] B. Poonen, *Average rank of elliptic curves [after Manjul Bhargava and Arul Shankar]*, Séminaire Bourbaki, Janvier 2012 / revised June 13, 64ème année, 2011-2012, no. 1049. [1](#), [1](#), [??](#), [4](#), [7](#), [7](#)
- [23] B. Poonen, E. Rains, *Random maximal isotropic subspaces and Selmer groups*, J. Amer. Math. Soc. 25 (2012), 245-269.
- [24] J. H. Silverman, *The arithmetic of elliptic curves*, Springer-Verlag, 2nd Edition, New York, 2009. [4](#)
- [25] P. Swinnerton-Dyer, *The effect of twisting on the 2-Selmer group*, Math. Proc. Cambridge Philos. Soc. 145 (2008), no. 3, 513-526. [4](#)
- [26] The SageMath Developers, Sage Mathematics Software (Version 6.9), 2015, <http://www.sagemath.org> [3.2](#)
- [27] M. Watkins, *Some heuristics about elliptic curves*, Experiment. Math. 17 (2008), no. 1, 105-125. [1](#), [6](#)

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CONNECTICUT, STORRS, CT 06269, USA

E-mail address: [alvaro.lozano-robledo@uconn.edu](mailto:alvaro.lozano-robledo@uconn.edu)

URL: <http://alozano.clas.uconn.edu/>